Appendices

A Visualizing the Set of Confusion Matrices

To clarify the geometry of the feasible set, we visualize one instance of the set of confusion matrices $\mathcal{C}$ using the dual representation of the supporting hyperplanes. This contains the following steps.

1. *Population Model:* We assume a joint probability for $\mathcal{X} = [-1, 1]$ and $\mathcal{Y} = \{0, 1\}$ given by

$$f_X = \mathbb{U}[-1, 1] \quad \text{and} \quad \eta(x) = \frac{1}{1 + e^{ax}}, \tag{12}$$

where $\mathbb{U}[-1, 1]$ is the uniform distribution on $[-1, 1]$ and $a > 0$ is a parameter controlling the degree of noise in the labels. If a is large, then with high probability, the true label is 1 on [-1, 0] and 0 on [0, 1]. On the contrary, if a is small, then there are no separable regions and the classes are mixed in $[-1, 1]$.

Furthermore, the integral $\int_{-1}^{1} \frac{1}{1+e^{ax}} dx = 1$ for $a \in \mathbb{R}$ implying $\mathbb{P}(Y = 1) = \zeta = \frac{1}{2} \; \forall \, a \in \mathbb{R}$.

2. *Generate Hyperplanes:* Take $\theta \in [0, 2\pi]$ and set $\mathbf{m} = (m_{11}, m_{00}) = (\cos\theta, \sin\theta)$. Let us denote x' as the point where the probability of positive class $\eta(x)$ is equal to the optimal threshold of Proposition 1. Solving for x in the equation $1/(1 + e^{ax}) = m_{00}/(m_{00} + m_{11})$ gives us

$$x' = \Pi_{[-1,1]}\big\{\tfrac{1}{a}\ln\big(\tfrac{m_{11}}{m_{00}}\big)\big\}, \tag{13}$$

where $\Pi_{[-1,1]}\{z\}$ is the projection of z on the interval $[-1, 1]$. If $m_{11} + m_{00} \geq 0$, then the Bayes classifier $\overline{h}$ predicts class 1 on the region $[-1, x']$ and 0 on the remaining region. If $m_{11}+m_{00} < 0$, $\overline{h}$ does the opposite. Using the fact that $Y|X$ and $\overline{h}|X$ are independent, we have that

(a) if $m_{11} + m_{00} \geq 0$, then

$$\overline{TP}_{\mathbf{m}} = \frac{1}{2}\int_{-1}^{x'} \tfrac{1}{1+e^{ax}} dx, \qquad \overline{TN}_{\mathbf{m}} = \tfrac{1}{2}\int_{x'}^{1} \tfrac{e^{ax}}{1+e^{ax}} dx.$$

(b) if $m_{11} + m_{00} < 0$, then

$$\overline{TP}_{\mathbf{m}} = \frac{1}{2}\int_{x'}^{1} \tfrac{1}{1+e^{ax}} dx, \qquad \overline{TN}_{\mathbf{m}} = \tfrac{1}{2}\int_{-1}^{x'} \tfrac{e^{ax}}{1+e^{ax}} dx.$$

Now, we can obtain the hyperplane as defined in (8) for each θ. We sample around thousand $\theta's \in [0, 2\pi]$ randomly, obtain the hyperplanes following the above process, and plot them.

The sets of feasible confusion matrices $\mathcal{C}$'s for $a = 0.5, 1, 2, 5, 10$, and 50 are shown in Figure 5. The middle white region is $\mathcal{C}$: the intersection of the half-spaces associated with its supporting hyperplanes. The curve on the right corresponds to the confusion matrices on the upper boundary $\partial\mathcal{C}_+$. Similarly, the curve on the left corresponds to the confusion matrices on the lower boundary $\partial\mathcal{C}_-$. Points $(\zeta, 0) = (\frac{1}{2}, 0)$ and $(0, 1 - \zeta) = (0, \frac{1}{2})$ are the two vertices. The geometry is 180°rotationally symmetric around the point $(\frac{1}{4}, \frac{1}{4})$.

Notice that as we increase the separability of the two classes via a, all the points in $[0, \zeta] \times [0, 1 - \zeta]$ becomes feasible. In other words, if the data is completely separable, then the corners on the top-right and the bottom left are achievable. If the data is 'inseparable', then the feasible set contains only the diagonal line joining $(0, \frac{1}{2})$ and $(\frac{1}{2}, 0)$, which passes through $(\frac{1}{4}, \frac{1}{4})$.

B Proofs

Lemma 4. *The feasible set of confusion matrices $\mathcal{C}$ has the following properties:*

(i). For all $(TP, TN) \in \mathcal{C}$, $0 \leq TP \leq \zeta$, and $0 \leq TN \leq 1 - \zeta$.

(ii). $(\zeta, 0) \in \mathcal{C}$ and $(0, 1 - \zeta) \in \mathcal{C}$.

(iii). For all $(TP, TN) \in \mathcal{C}$, $(\zeta - TP, 1 - \zeta - TN) \in \mathcal{C}$.

(iv). $\mathcal{C}$ is convex.

(v). $\mathcal{C}$ has a supporting hyperplane associated to every normal vector.

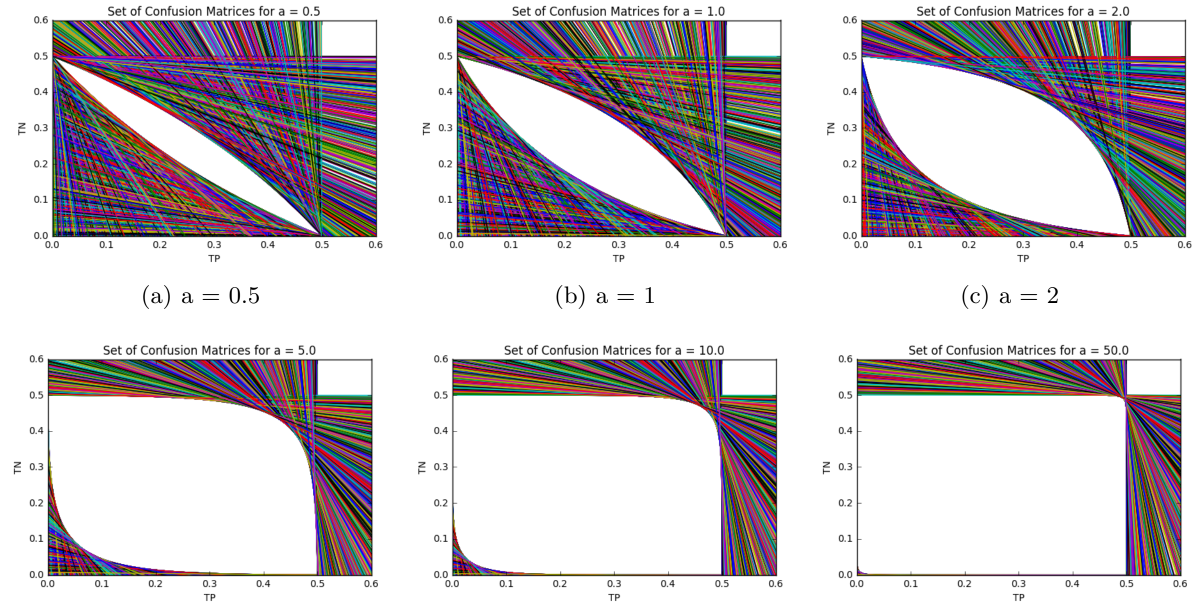
(vi). Any supporting hyperplane with positive slope is tangent to $\mathcal{C}$ at $(0, 1 - \zeta)$ or $(\zeta, 0)$.

Proof. We prove the statements as follows:

(i). $0 \leq \mathbb{P}[h = Y = 1] \leq \mathbb{P}[Y = 1] = \zeta$, and similarly, $0 \leq \mathbb{P}[h = Y = 0] \leq \mathbb{P}[Y = 0] = 1 - \zeta$.

(ii). If h is the trivial classifier which always predicts 1, then $TP(h) = \Pr[h = Y = 1] = \Pr[Y = 1] = \zeta$, and $TN(h) = 0$. This means that $(\zeta, 0) \in \mathcal{C}$. Similarly, if h is the classifier which always predicts 0, then $TP(h) = \Pr[h = Y = 1] = 0$, and $TN(h) = \Pr[h = Y = 0] = \Pr[Y = 0] = 1 - \zeta$. Therefore, $(0, 1 - \zeta) \in \mathcal{C}$.

(iii). Let h be a classifier such that $TP(h) = TP$, $TN(h) = TN$. Now, consider the classifier $1 - h$



(a) a = 0.5 (b) a = 1 (c) a = 2

(d) a = 5 (e) a = 10 (f) a = 50

Fig. 5: Supporting hyperplanes and associated set of feasible confusion matrices for exponential model (12) with $a = 0.5, 1, 2, 5, 10$ and 50. The middle white region is $\mathcal{C}$: the intersection of half-spaces associated with its supporting hyperplanes.

(which predicts exactly the opposite of h). We have that

$$\begin{aligned} TP(1-h) &= \mathbb{P}[(1-h) = Y = 1] \\ &= \mathbb{P}[Y=1] - \mathbb{P}[h = Y = 1] \\ &= \zeta - TP(h). \end{aligned}$$

A similar argument gives

$$TN(1-h) = 1 - \zeta - TN(h).$$

(iv). Consider any two confusion matrices (TP_1, TN_1), $(TP_2, TN_2) \in \mathcal{C}$, attained by the classifiers $h_1, h_2 \in \mathcal{H}$, respectively. Let $0 \leq \lambda \leq 1$. Define a classifier h' which predicts the output from the classifier h_1 with probability λ and predicts the output of the classifier h_2 with probability $1 - \lambda$. Then,

$$\begin{aligned} TP(h') &= \mathbb{P}[h' = Y = 1] \\ &= \mathbb{P}[h_1 = Y = 1 | h = h_1]\mathbb{P}[h = h_1] \\ &+ \mathbb{P}[h_2 = Y = 1 | h = h_2]\mathbb{P}[h = h_2] \\ &= \lambda TP(h_1) + (1-\lambda) TP(h_2). \end{aligned}$$

A similar argument gives the convex combination for TN. Thus, $\lambda(TP(h_1), TN(h_1)) + (1 - \lambda)(TP(h_2), TN(h_2)) \in \mathcal{C}$ and hence, $\mathcal{C}$ is convex.

(v). This follows from convexity (iv) and boundedness (i).

(vi). For any bounded, convex region in $[0,\zeta]\times[0,1-\zeta]$ which contains the points $(0,\zeta)$ and $(0,1-\zeta)$, it is true that any positively sloped supporting hyperplane will be tangent to $(0,\zeta)$ or $(0,1-\zeta)$.

□

Lemma 5. *The boundary of $\mathcal{C}$ is exactly the confusion matrices of estimators of the form $\lambda\mathbb{1}[\eta(x) \geq t] + (1 - \lambda)\mathbb{1}[\eta(x) > t]$ and $\lambda\mathbb{1}[\eta(x) < t] + (1-\lambda)\mathbb{1}[\eta(x) \leq t]$ for some $\lambda, t \in [0,1]$.*

Proof. To prove that the boundary is attained by estimators of these forms, consider solving the problem under the constraint $\mathbb{P}[h = 1] = c$. We have $\mathbb{P}[h=1] = TP + FP$, and $\zeta = \mathbb{P}[Y=1] = TP + FN$, so we get

$$TP - TN = c + \zeta - TP - TN - FP - FN = c + \zeta - 1,$$

which is a constant. Note that no confusion matrix has two values of $TP - TN$. This effectively partitions $\mathcal{C}$, since all confusion matrices are attained by varying c from 0 to 1. Furthermore, since $A := TN = TP - c - \zeta + 1$ is an affine space (a line in tp-tn coordinate system), $\mathcal{C} \cap A$ has at least one endpoint, because A would pass through the box $[\zeta, 0] \times [0, 1-\zeta]$ and has at most two endpoints due to convexity and boundedness of $\mathcal{C}$. Since A is a line with positive slope, $\mathcal{C} \cap A$ is a single point only when A is tangent to $\mathcal{C}$ at $(0, 1-\zeta)$ or $(\zeta, 0)$, from Lemma 4, part (vi).

Since the affine space A has positive slope, we claim that the two endpoints are attained by maximizing or minimizing $TP(h)$ subject to $\Pr[h = 1] = c$. It remains to show that this happens for estimators of the form $h^\lambda_{t+} := \lambda \mathbb{1}[\eta(x) \geq t] + (1-\lambda)\mathbb{1}[\eta(x) > t]$ and $h^\lambda_{t-} := \lambda \mathbb{1}[\eta(x) < t] + (1-\lambda)\mathbb{1}[\eta(x) \leq t]$, respectively.

Let h be any estimator, and recall

$$TP(h) := \int_{\mathcal{X}} \eta(x)\mathbb{P}[h = 1|X = x]\,\mathrm{d}f_X.$$

It should be clear that under a constraint $\mathbb{P}[h = 1] = c$, the optimal choice of h puts all the weight onto the larger values of η. One can begin by classifying those X into the positive class where $n(X)$ is maximum, until one exhausts the budget of c. Let t be such that $\mathbb{P}[h^0_{t+} = 1] \leq c \leq \mathbb{P}[h^1_{t+} = 1]$, and let $\lambda \in [0, 1]$ be chosen such that $\mathbb{P}[h^\lambda_{t+} = 1] = c$, then h^λ_{t+} must maximize $TP(h)$ subject to $\mathbb{P}[h = 1] = c$.

A similar argument shows that all TP-minimizing boundary points are attained by the h_{t-}'s. □

Remark 1. *Under Assumption 1,* $\mathbb{1}[\eta(x) > t] = \mathbb{1}[\eta(x) \geq t]$ *and* $\mathbb{1}[\eta(x) < t] = \mathbb{1}[\eta(x) \leq t]$. *Thus, the boundary of* $\mathcal{C}$ *is the confusion matrices of estimators of the form* $\mathbb{1}[\eta(x) \geq t]$ *and* $\mathbb{1}[\eta(x) \leq t]$ *for some* $t \in [0, 1]$.

Proof of Proposition 1. "Let $\phi \in \varphi_{LPM}$*, then*

$$\overline{h}(x) = \left\{ \begin{array}{ll} \mathbb{1}[\eta(x) \geq \frac{m_{00}}{m_{11}+m_{00}}], & m_{11} + m_{00} \geq 0 \\ \mathbb{1}[\frac{m_{00}}{m_{11}+m_{00}} \geq \eta(x)], & o.w. \end{array} \right\}$$

is a Bayes optimal classifier w.r.t ϕ. *Further, the inverse Bayes classifier is given by* $\underline{h} = 1 - \overline{h}$*."*

Note, we are maximizing a linear function on a convex set. There are 6 cases to consider:

1. If the signs of m_{11} and m_{00} differ, the maximum is attained either at $(0, 1-\zeta)$ or $(\zeta, 0)$, as per Lemma 4, part (vi). Which of the two is optimum depends on whether $|m_{11}| \geq |m_{00}|$, i.e. on the sign of $m_{11}+m_{00}$. It should be easy to check that in all 4 possible cases, the statement holds, noting that in all 4 cases, $0 \leq m_{00}/(m_{11} + m_{00}) \leq 1$.

2. If $m_{11}, m_{00} \geq 0$, then the maximum is attained on $\partial\mathcal{C}_+$, and the proof below gives the desired result.

   We know, from Lemma 5, that $\overline{h}$ must be of the form $\mathbb{1}[\eta(x) \geq t]$ for some t. It suffices to find t. Thus, we wish to maximize $m_{11}TP(h_t) + m_{00}TN(h_t)$. Now, let $Z := \eta(X)$ be the random variable obtained by evaluating η at random X. Under Assumption 1, $df_X = df_Z$ and we have that

   $$TP(h_t) \;=\; \int_{x:\eta(x)\geq t} \eta(x)\,\mathrm{d}f_X \;=\; \int_t^1 z\,\mathrm{d}f_Z$$

   Similarly, $TN(h_t) = \int_0^t (1-z)\,\mathrm{d}f_Z$. Therefore,

   $$\begin{aligned}&\tfrac{\partial}{\partial t}\big(m_{11}TP(h_t) + m_{00}TN(h_t)\big) \\ &\qquad = -m_{11}tf_Z(t) + \cdot m_{00}(1-t)f_Z(t).\end{aligned}$$

   So, the critical point is attained at $t = m_{00}/(m_{11} + m_{00})$, as desired. A similar argument gives the converse result for $m_{11} + m_{00} < 0$.

3. if $m_{11}, m_{00} < 0$, then the maximum is attained on $\partial\mathcal{C}_-$, and an argument identical to the proof above gives the desired result.

□

Proof of Proposition 2. "The set of confusion matrices $\mathcal{C}$ *is convex, closed, contained in the rectangle* $[0, \zeta] \times [0, 1-\zeta]$ *(bounded), and* $180°$ *rotationally symmetric around the center-point* $(\frac{\zeta}{2}, \frac{1-\zeta}{2})$*. Under Assumption 1,* $(0, 1-\zeta)$ *and* $(\zeta, 0)$ *are the only vertices of* $\mathcal{C}$*, and* $\mathcal{C}$ *is strictly convex. Thus, any supporting hyperplane of* $\mathcal{C}$ *is tangent at only one point."*

That $\mathcal{C}$ is convex and bounded is already proven in Lemma 4. To see that $\mathcal{C}$ is closed, note that, from Lemma 5, every boundary point is attained. From Lemma 4, part (iii), it follows that $\mathcal{C}$ is $180°$ rotationally symmetric around the point $(\frac{\zeta}{2}, \frac{1-\zeta}{2})$.

Further, recall every boundary point of $\mathcal{C}$ can be attained by a thresholding estimator. By the discussion in Section 3, every boundary point is the optimal classifier for some linear performance metric, and the vector defining this linear metric is exactly the normal vector of the supporting hyperplane at the boundary point.

A vertex exists if (and only if) some point is supported by more than one tangent hyperplane in two dimensional space. This means it is optimal for more than one linear metric. Clearly, all the hyperplanes corresponding to the slope of the metrics where m_{11} and m_{00} are of opposite sign (i.e. hyperplanes with positive slope) support either $(\zeta, 0)$ or $(0, 1-\zeta)$. So, there are at least two supporting hyperplanes at these points, which make them the vertices. Now, it remains to show that there are no other vertices for the set $\mathcal{C}$.

Now consider the case when the slopes of the hyperplanes are negative, i.e. m_{11} and m_{00} have the same sign for the corresponding linear metrics. We know from Proposition 1 that optimal classifiers for linear metrics are threshold classifiers. Therefore there exist more than one threshold classifier of the form $h_t = \mathbb{1}[\eta(x) \geq t]$ with the same confusion matrix. Let's call them h_{t_1} and h_{t_2} for the two thresholds $t_1, t_2 \in [0,1]$. This means that $\int_{x:\eta(x)\geq t_1} \eta(x) df_X = \int_{x:\eta(x)\geq t_2} \eta(x) df_X$. Hence, there are multiple values of η which are never attained! This contradicts that g is strictly decreasing. Therefore, there are no vertices other than $(\zeta, 0)$ or $(0, 1-\zeta)$ in $\mathcal{C}$.

Now, we show that no supporting hyperplane is tangent at multiple points (i.e., there no flat regions on the boundary). If suppose there is a hyperplane which supports two points on the boundary. Then there exist two threshold classifiers with arbitrarily close threshold values, but confusion matrices that are well-separated. Therefore, there must exist some value of η which exists with non-zero probability, contradicting the continuity of g. By the discussion above, we conclude that under Assumption 1, every supporting hyperplane to the convext set $\mathcal{C}$ is tangent to only one point. This makes the set $\mathcal{C}$ strictly convex. $\square$

Proof of Lemma 1. "Let $\rho^+ : [0,1] \to \partial\mathcal{C}_+$, $\rho^- : [0,1] \to \partial\mathcal{C}_-$ be continuous, bijective, parametrizations of the upper and lower boundary, respectively. Let $\phi : \mathcal{C} \to \mathbb{R}$ be a quasiconcave function, and $\psi : \mathcal{C} \to \mathbb{R}$ be a quasiconvex function, which are monotone increasing in both TP and TN. Then the composition $\phi \circ \rho^+ : [0,1] \to \mathbb{R}$ is quasiconcave (and therefore unimodal) on the interval $[0,1]$, and $\psi \circ \rho^- : [0,1] \to \mathbb{R}$ is quasiconvex (and therefore unimodal) on the interval $[0,1]$."

We will prove the result for $\phi \circ \rho^+$ on $\partial\mathcal{C}^+$, and the argument for $\psi \circ \rho^-$ on $\partial\mathcal{C}^+$ is essentially the same. For simplicity, we drop the + symbols in the notation. Recall that a function is quasiconcave if and only if its superlevel sets are convex.

It is given that ϕ is quasiconcave. Let S be some superlevel set of ϕ. We first want to show that for any $r < s < t$, if $\rho(r) \in S$ and $\rho(t) \in S$, then $\rho(s) \in S$. Since ρ is a continuous bijection, due to the geometry of $\mathcal{C}$ (Lemma 4 and Proposition 2), we must have — without loss of generality — $TP(\rho(r)) < TP(\rho(s)) < TP(\rho(t))$, and $TN(\rho(r)) > TN(\rho(s)) > TN(\rho(t))$. (otherwise swap r and t). Since the set $\mathcal{C}$ is strictly convex and the image of ρ is $\partial\mathcal{C}$, then $\rho(s)$ must dominate (componentwise) a point in the convex combination of $\rho(r)$ and $\rho(t)$. Say that point is z. Since ϕ is monotone increasing, then $x \in S \implies y \in S$ for all $y \geq x$ componentwise. Thereofore, $\phi(\rho(s)) \geq \phi(z)$. Since, S is convex, $z \in S$ and, due to the argument above, $\rho(s) \in S$.

This implies that $\rho^{-1}(\partial\mathcal{C} \cap S)$ is an interval, and is therefore convex. Thus, the superlevel sets of $\phi \circ \rho$ are convex, so it is quasiconcave, as desired. This implies unimodaltiy as a function over the real line which has more than one local maximum can not be quasiconcave (consider the super-level set for some value slightly less than the lowest of the two peaks). $\square$

Proof of Proposition 3. "Sufficient conditions for $\phi \in \varphi_{LFPM}$ to be bounded in $[0,1]$ and simultaneously monotonically increasing in TP and TN are: $p_{11}, p_{00} \geq 0$, $p_{11} \geq q_{11}$, $p_{00} \geq q_{00}$, $q_0 = (p_{11} - q_{11})\zeta + (p_{00} - q_{00})(1-\zeta) + p_0$, $p_0 = 0$, and $p_{11} + p_{00} = 1$ (Conditions in Assumpotion 2). WLOG, we can take both the numerator and denominator to be positive."

For this proof, we denote TP and TN as C_{11} and C_{00}, respectively. Let us take a linear-fractional metric

$$\phi(C) = \frac{p_{11}C_{11} + p_{00}C_{00} + p_0}{q_{11}C_{11} + q_{00}C_{00} + q_0} \tag{14}$$

where $p_{11}, q_{11}, p_{00}, q_{00}$ are not zero simultaneously. We want $\phi(C)$ to be monotonic in TP, TN and bounded. If for any $C \in \mathcal{C}$, $\phi(C) < 0$, we can add a large positive constant such that $\phi(C) \geq 0$, and still the metric would remain linear fractional. So, it is sufficient to assume $\phi(C) \geq 0$. Furthermore, boundedness of ϕ implies $\phi(C) \in [0, D]$, for some $\mathbb{R} \ni D \geq 0$. Therefore, we may divide $\phi(C)$ by D so that $\phi(C) \in [0,1]$ for all $C \in \mathcal{C}$. Still, the metric is linear fractional and $\phi(C) \in [0,1]$.

Taking derivative of $\phi(C)$ w.r.t. C_{11}.

$$\begin{aligned}\frac{\partial\phi(C)}{\partial C_{11}} &= \frac{p_{11}}{q_{11}C_{11} + q_{00}C_{00} + q_0} \\ &\quad - \frac{q_{11}(p_{11}C_{11} + p_{00}C_{00} + p_0)}{(q_{11}C_{11} + q_{00}C_{00} + q_0)^2} \geq 0\end{aligned}$$

$$\Rightarrow p_{11}(q_{11}C_{11} + q_{00}C_{00} + q_0) \geq q_{11}(p_{11}C_{11} + p_{00}C_{00} + p_0)$$

If denominator is positive then the numerator is positive as well.

- Case 1: The denominator $q_{11}C_{11} + q_{00}C_{00} + q_0 \geq 0$.
  - Case (a) $q_{11} > 0$.

    $$\begin{aligned}&\Rightarrow p_{11} \geq q_{11}\phi(C) \\ &\Rightarrow p_{11} \geq q_{11} \sup_{C\in\mathcal{C}} \phi(C) \\ &\Rightarrow p_{11} \geq q_{11}\overline{\tau} \qquad \text{(Necessary Condition)}\end{aligned}$$

    We are considering sufficient condition, which means $\overline{\tau}$ can vary from $[0,1]$. Hence, a sufficient condition for monotonicity in C_{11} is $p_{11} \geq q_{11}$. Furthermore, $p_{11} \geq 0$ as well.

  - Case (b) $q_{11} < 0$.

$$\Rightarrow p_{11} \geq q_{11}\overline{\tau}$$

Since $q_{11} < 0$ and $\overline{\tau} \in [0,1]$, sufficient condition is $p_{11} \geq 0$. So, in this case as well we have that

$$p_{11} \geq q_{11}, \;\; p_{11} \geq 0.$$

  - Case(c) $q_{11} = 0$.

$$\Rightarrow p_{11} \geq 0$$

We again have $p_{11} \geq q_{11}$ and $p_{11} \geq 0$ as sufficient conditions.
A similar case holds for C_{00}, implying $p_{00} \geq q_{00}$ and $p_{00} \geq 0$.

- Case 2: The denominator $q_{11}C_{11} + q_{00}C_{00} + q_0$ is negative.

$$p_{11} \leq q_{11}\Big(\frac{p_{11}C_{11} + p_{00}C_{00} + p_0}{q_{11}C_{11} + q_{00}C_{00} + q_0}\Big)$$
$$\Rightarrow p_{11} \leq q_{11}\overline{\tau}$$

  - Case(a) If $q_{11} > 0$. So, we have $p_{11} \leq q_{11}$ and $p_{11} \leq 0$ as sufficient condition.
  - Case(b) If $q_{11} < 0, \Rightarrow p_{11} \leq q_{11}$. So, we have $q_{11} < 0, \Rightarrow p_{11} < 0$ as sufficient condition.
  - Case(c) If $q_{11} = 0, \Rightarrow p_{11} \leq 0$ and $p_{11} \leq q_{11}$ as sufficient condition.
    So in all the cases we have that

$$p_{11} \leq q_{11} \text{ and } p_{11} \leq 0$$

as the sufficient conditions. A similar case holds for C_{00} resulting in $p_{00} \leq q_{00}$ and $p_{00} \leq 0$.

Suppose the points where denominator is positive is $\mathcal{C}^+ \subseteq \mathcal{C}$. Suppose the points where denominator is negative is $\mathcal{C}^- \subseteq \mathcal{C}$. For gradient to be non-negative at points belonging to $\mathcal{C}^+$, the sufficient condition is

$$p_{11} \geq q_{11} \text{ and } p_{11} \geq 0$$
$$p_{00} \geq q_{00} \text{ and } p_{00} \geq 0$$

For gradient to be non-negative at points belonging to $\mathcal{C}^-$, the sufficient condition is

$$p_{11} \leq q_{11} \text{ and } p_{11} \leq 0$$
$$p_{00} \leq q_{00} \text{ and } p_{00} \leq 0$$

If $\mathcal{C}_+$ and $\mathcal{C}_-$ are not empty sets, then the gradient is non-negative only when $p_{11}, p_{00} = 0$ and $q_{11}, q_{00} = 0$. This is not possible by the definition described in (14). Hence, one of $\mathcal{C}_+$ or $\mathcal{C}_-$ should be empty. WLOG, we assume $\mathcal{C}_-$ is empty and conclude that $\mathcal{C}_+ = \mathcal{C}$.
An immediate consequence of this is, WLOG, we can take both the numerator and the denominator to be positive, and the sufficient conditions for monotonicity are as follows:

$$p_{11} \geq q_{11} \text{ and } p_{11} \geq 0$$
$$p_{00} \geq q_{00} \text{ and } p_{00} \geq 0$$

Now, let us take a point in the feasible space $(\zeta, 0)$. We know that

$$\begin{aligned}
\phi((\zeta,0)) = \frac{p_{11}\zeta + p_0}{q_{11}\zeta + q_0} &\leq \overline{\tau} \\
&\Rightarrow p_{11}\zeta + p_0 \leq \overline{\tau}(q_{11}\zeta + q_0) \\
&\Rightarrow (p_{11} - \overline{\tau}q_{11})\zeta + (p_0 - \overline{\tau}q_0) \leq 0 \\
&\Rightarrow (p_0 - \overline{\tau}q_0) \leq -\underbrace{(p_{11} - \overline{\tau}q_{11})}_{\text{positive}}\underbrace{\zeta}_{\text{positive}} \\
&\Rightarrow (p_0 - \overline{\tau}q_0) \leq 0. \qquad (15)
\end{aligned}$$

Metric being bounded in $[0,1]$ gives us

$$\frac{p_{11}C_{11} + p_{00}C_{00} + p_0}{q_{11}C_{11} + q_{00}C_{00} + q_0} \leq 1$$
$$\Rightarrow p_{11}C_{11} + p_{00}C_{00} + p_0 \leq q_{11}C_{11} + q_{00}C_{00} + q_0$$

$$\Rightarrow q_0 \geq (p_{11} - q_{11})c_{11} + (p_{00} - q_{00})c_{00} + p_0 \qquad \forall C \in \mathcal{C}.$$

Hence, a sufficient condition is

$$q_0 = (p_{11} - q_{11})\zeta + (p_{00} - q_{00})(1 - \zeta) + p_0.$$

Equation (15), which we derived from monotonicity, implies that

- Case (a) $q_0 \geq 0, \Rightarrow p_0 \leq 0$ as a sufficient condition.
- Case (b) $q_0 \leq 0, \Rightarrow p_0 \leq q_0 \leq 0$ as a sufficient condition.

Since the numerator is positive for all $C \in \mathcal{C}$ and $p_{11}, p_{00} \geq 0$, a sufficient condition for p_0 is $p_0 = 0$.

Finally, a monotonic, bounded in $[0,1]$, linear fractional metric is defined by

$$\phi(C) = \frac{p_{11}c_{11} + p_{00}c_{00} + p_0}{q_{11}c_{11} + q_{00}c_{00} + q_0},$$

where $p_{11} \geq q_{11}, p_{11} \geq 0, p_{00} \geq q_{00}, p_{00} \geq 0, q_0 = (p_{11} - q_{11})\zeta + (p_{00} - q_{00})(1 - \zeta) + p_0, p_0 = 0$, and p_{11}, q_{11}, p_{00}, and q_{00} are not simulataneously zero. Further, we can divide the numerator and denominator with $p_{11} + p_{00}$ without changing the metric ϕ and the above sufficient conditions. Therefore, for elicitation purposes, we can take $p_{11} + p_{00} = 1$. $\square$

Proof of Proposition 4. *"Under Assumption 2, knowing p'_{11} solves the system of equations (9) as follows:*

$$p'_{00} = 1 - p'_{11},\ q'_0 = \overline{C}_0 \frac{P'}{Q'},$$
$$q'_{11} = (p'_{11} - \overline{m}_{11})\frac{P'}{Q'},\ q'_{00} = (p'_{00} - \overline{m}_{00})\frac{P'}{Q'}, \tag{16}$$

where $P' = p'_{11}\zeta + p'_{00}(1-\zeta)$ and $Q' = P' + \overline{C}_0 - \overline{m}_{11}\zeta - \overline{m}_{00}(1-\zeta)$. Thus, it elicits the LFPM."

For this proof as well, we use $TP = C_{11}$ and $TN = C_{00}$. Since the linear fractional matrix is monotonically increasing in C_{11} and C_{00}, it is maximized at the upper boundary $\partial\mathcal{C}_+$. Hence $m_{11} \geq 0$ and $m_{00} \geq 0$. So, after running Algorithm 1, we get a hyperplane such that

$$p_{11} - \tau q_{11} = \alpha m_{11}, \quad p_{00} - \tau q_{00} = \alpha m_{00},$$
$$p_0 - \tau q_0 = -\alpha \underbrace{(m_{11}C^*_{11} + m_{00}C^*_{00})}_{=:C_0}. \tag{17}$$

Since $p_{11} - \bar{\tau} q_{11} \geq 0$ and $m_{11} \geq 0$, $\Rightarrow \alpha \geq 0$. As discussed in the main paper, we avoid the case when $\alpha = 0$. Therefore, we have that $\alpha > 0$.

Equation (17) implies that

$$\frac{p_{11}}{\alpha} - \frac{\tau q_{11}}{\alpha} = m_{11}, \quad \frac{p_{00}}{\alpha} - \frac{\tau q_{00}}{\alpha} = m_{00},$$
$$\frac{p_0}{\alpha} - \frac{\tau q_0}{\alpha} = -C_0.$$

Assume $p'_{11} = \frac{p_{11}}{\alpha}, p'_{00} = \frac{p_{00}}{\alpha}$, $q'_{11} = \frac{q_{11}}{\alpha}$, $q'_{00} = \frac{q_{00}}{\alpha}$, $p'_0 = \frac{p_0}{\alpha}$, $q'_0 = \frac{q_0}{\alpha}$. Then, the above system of equations turns into

$$p'_{11} - \bar{\tau} q'_{11} = m_{11}, \quad p'_{00} - \bar{\tau} q'_{00} = m_{00},$$
$$p'_0 - \bar{\tau} q'_0 = -C_0.$$

A ϕ' metric defined by the $p'_{11}, p'_{00}, q'_{11}, q'_{00}, q'_0$ is monotonic, bounded in $[0,1]$, and satisfies all the sufficient conditions of Assumptions 2, i.e.,

$$p'_{11} \geq q'_{11}\ ,\ p'_{00} \geq q'_{11},\ p'_{11} \geq 0\ ,\ p'_{00} \geq 0,$$
$$q'_0 = (p'_{11} - q_{11})\pi + (p'_{00} - q'_{00})\pi + p'_0,\ p'_0 = 0.$$

As discussed in the main paper, solving the above system does not harm the elicitation task. For simplicity, replacing the " $'$ " notation with the normal one, we have that

$$p_{11} - \bar{\tau} q_{11} = m_{11}, \quad p_{00} - \bar{\tau} q_{00} = m_{00},$$
$$p_0 - \bar{\tau} q_0 = -C_0$$

From last equation, we have that $\bar{\tau} = \frac{C_0 + p_0}{q_0}$. Putting it in the rest gives us

$$q_0 p_{11} - (C_0 + p_0) q_{11} = m_{11} q_0,$$
$$q_0 p_{00} - (C_0 + p_0) q_{00} = m_{00} q_0.$$

We already have

$$q_0 = (p_{11} - q_{11})\zeta + (p_{00} - q_{00})(1-\zeta) + p_0$$
$$\Rightarrow q_{11} = \frac{p_{00}(1-\zeta) - q_{00}(1-\zeta) + p_{11}\zeta - q_0 + p_0}{\zeta},$$

which further gives us

$$q_0 = \frac{(C_0 + p_0)[p_{00}(1-\zeta) + p_{11}\zeta + p_0]}{p_{11}\zeta + p_{00}(1-\zeta) + p_0 + C_0 - m_{11}\zeta - m_{00}(1-\zeta)},$$
$$q_{00} = \frac{(p_{00} - m_{00})[p_{00}(1-\zeta) + p_{11}\zeta + p_0]}{p_{11}\zeta + p_{00}(1-\zeta) + p_0 + C_0 - m_{11}\zeta - m_{00}(1-\zeta)},$$
$$q_{11} = \frac{(p_{11} - m_{11})[p_{00}(1-\zeta) + p_{11}\zeta + p_0]}{p_{11}\zeta + p_{00}(1-\zeta) + p_0 + C_0 - m_{11}\zeta - m_{00}(1-\zeta)}.$$

Define

$$P := p_{00}(1-\zeta) + p_{11}\zeta + p_0,$$
$$Q := P + C_0 - m_{11}\zeta - m_{00}(1-\zeta).$$

Hence,

$$q_0 = (C_0 + p_0)\frac{P}{Q}, \quad q_{11} = (p_{11} - m_{11})\frac{P}{Q},$$
$$q_{00} = (p_{00} - m_{00})\frac{P}{Q}.$$

Now using sufficient conditions, we have $p_0 = 0$. The final solution is the following:

$$q_0 = C_0 \frac{P}{Q}, \quad q_{11} = (p_{11} - m_{11})\frac{P}{Q},$$
$$q_{00} = (p_{00} - m_{00})\frac{P}{Q},$$

where $P := p_{11}\zeta + p_{00}(1-\zeta)$ and $Q := P + C_0 - m_{11}\zeta - m_{00}(1-\zeta)$. We have taken $p_{11} + p_{00} = 1$, but the original $p'_{11} + p'_{00} = \frac{1}{\alpha}$. Therefore, we learn $\hat{\phi}(C)$ such that such that $\hat{\phi}(C) = \alpha\phi(C)$. □

Corollary 1. *For F_β-measure, where β is unknown, Algorithm 1 elicits the true performance metric up to a constant in $O(\log(\frac{1}{\epsilon}))$ queries to the oracle.*

Proof. Algorithm 1 gives us the supporting hyperplane, the trade-off, and the Bayes confusion matrix. If we know p_{11}, then we can use Proposition 4 to compute the other coefficients. In F_β-measure, $p_{11} = 1$, and we do not require Algorithms 2 and 3. □

Proof of Theorem 1. *"Given $\epsilon, \epsilon_\Omega \geq 0$ and a 1-Lipschitz metric ϕ that is monotonically increasing in TP, TN. If it is quasiconcave (quasiconvex) then*

Algorithm 1 (Algorithm 2) finds an approximate maximizer $\overline{C}$ (minimizer $\underline{C}$). Furthemore, (i) the algorithm returns the supporting hyperplane at that point, (ii) the value of ϕ at that point is within $O(\sqrt{\epsilon_\Omega} + \epsilon)$ of the optimum, and (iii) the number of queries is $O(\log \frac{1}{\epsilon})$."

(i) As a direct consequence of our representation of the points on the boundary via their supporting hyperplanes (Section 3.1), when we search for the maximizer (mimimizer), we also get the associated supporting hyperplane as well.

(ii) By the nature of binary search, we are effectively narrowing our search interval around some target angle θ_0. Furthermore, since the oracle queries are correct unless the ϕ values are within ϵ_Ω, we must have $|\phi(C_{\overline{\theta}}) - \phi(C_{\theta_0})| < \epsilon_\Omega$, and we output θ' such that $|\theta_0 - \theta'| < \epsilon$. Now, we want to check the bound $|\phi(C_{\theta'}) - \phi(C_{\overline{\theta}})|$. In order to do that, we will also consider the threshold corresponding to the supporting hyperplanes at C_θ's, i.e. $\delta_\theta = {}^{\sin\theta}/_{\sin\theta + \cos\theta}$.

Notice that,

$$
\begin{aligned}
|\phi(C_{\overline{\theta}}) - \phi(C_{\theta'})| &= |\phi(C_{\overline{\theta}}) - \phi(C_{\theta_0}) \\
&\quad + \phi(C_{\theta_0}) - \phi(C_{\theta'})| \\
&\leq |\phi(C_{\overline{\theta}}) - \phi(C_{\theta_0})| \\
&\quad + |\phi(C_{\theta_0}) - \phi(C_{\theta'})| \qquad (18)
\end{aligned}
$$

The first term is bounded by ϵ_Ω due to the oracle assumption. For the bounds the second term, consider the following.

$$|TP(C_{\theta_0}) - TP(C_{\theta'})|$$

$$
\begin{aligned}
&= \left| \int_{x: \frac{sin\theta_0}{sin\theta_0 + cos\theta_0} \geq \eta(x) \geq \frac{sin\theta'}{sin\theta' + cos\theta'}} \eta(x)\, \mathrm{d}f_X \right| \\
&\leq \left| \int_{x: \frac{sin\theta_0}{sin\theta_0 + cos\theta_0} - \overline{\delta} \geq \eta(x) - \overline{\delta} \geq \frac{sin\theta'}{sin\theta' + cos\theta'} - \overline{\delta}} \mathrm{d}f_X \right| \\
&= \left| \int_{x: \frac{sin\theta_0}{sin\theta_0 + cos\theta_0} - \frac{sin\overline{\theta}}{sin\overline{\theta} + cos\overline{\theta}} \geq \eta(x) - \overline{\delta} \geq \frac{sin\theta'}{sin\theta' + cos\theta'} - \frac{sin\overline{\theta}}{sin\overline{\theta} + cos\overline{\theta}}} \mathrm{d}f_X \right| \\
&= \left| \int_{x: \frac{sin(\theta_0 - \overline{\theta})}{sin(\theta_0 + \overline{\theta}) + cos(\theta_0 - \overline{\theta})} \geq \eta(x) - \overline{\delta} \geq \frac{sin\theta'}{sin\theta' + cos\theta'} - \frac{sin\overline{\theta}}{sin\overline{\theta} + cos\overline{\theta}}} \mathrm{d}f_X \right|,
\end{aligned}
\qquad (19)
$$

where the inequality in the second step follows from the fact that $\eta(x) \leq 1$.

Recall that the left term in the integral limits is actually, $\delta_{\theta_0} - \delta_{\overline{\theta}}$. When $|\phi(C_{\delta_{\theta_0}}) - \phi(C_{\delta_{\overline{\theta}}})| < \epsilon_\Omega$, then we have $|\overline{\delta} - \delta_0| < \frac{2}{k_0}\sqrt{k_1 \epsilon_\Omega}$. The proof of this statement is given in the proof of Theorem 2 (proved later). Since sin is 1-Lipschitz, adding and subtracting $\sin\theta_0/(\sin\theta_0 + \cos\theta_0)$ in the right term of the integration limit gives us the minimum value of the right term to be $-\epsilon - \frac{2\sqrt{k_1 \epsilon_\Omega}}{k_0}$. This implies that the quantity in (19) is less than

$$
\begin{aligned}
&\mathbb{P}[\{(\eta(X) - \overline{\delta}) \leq \frac{2}{k_0}\sqrt{k_1 \epsilon_\Omega}\} \cap \\
&\qquad \{(\overline{\delta} - \eta(X)) \leq \epsilon + \frac{2}{k_0}\sqrt{k_1 \epsilon_\Omega}\}] \\
&\leq \mathbb{P}[(\overline{\delta} - \eta(X)) \leq \epsilon + \frac{2}{k_0}\sqrt{k_1 \epsilon_\Omega}] \\
&\leq \frac{2k_1}{k_0}\sqrt{k_1 \epsilon_\Omega} + k_1 \epsilon \quad \text{(by Assumption 4)} \qquad (20)
\end{aligned}
$$

As $\mathbb{P}(A \cap B) \leq min\{\mathbb{P}(A), \mathbb{P}(B)\}$, the inequality used in the second step is rather loose, but it shows the dependency on sufficiently small ϵ. It could be independent of the tolerance ϵ depending on the $\mathbb{P}(\eta(X) - \overline{\delta})$ or the sheer big value of ϵ. Nevertheless, a similar result applies to the true negative rate. Since ϕ is 1-Lipschitz, we have that $|\phi(C) - \phi(C')| \leq 1 \cdot \|C - C'\|$, but

$$\|C(\theta_0) - C(\theta')\|_\infty \leq \frac{2k_1}{k_0}\sqrt{k_1 \epsilon_\Omega} + k_1 \epsilon.$$

Hence, $|\phi(C_{\theta'}) - \phi(C_{\overline{\theta}})| \leq \sqrt{2}(\frac{2k_1}{k_0}\sqrt{k_1 \epsilon_\Omega} + k_1 \epsilon) + \epsilon_\Omega$. Since the metrics are in $[0, 1]$, $\epsilon_\Omega \in [0, 1]$. Therefore, $\sqrt{\epsilon_\Omega} \geq \epsilon_\Omega$. This gives us the desired result.

(iii) We needed only, for part (ii), that the interval of possible values of θ' be at most ϵ to the target angle θ_0. Ideally, this is obtained by making $\log_2(1/\epsilon)$ queries, but due to the region where oracle misreport its preferences, we can be off to the target angle θ_0 by more than ϵ. However, binary search will again put us back in the correct direction, once we leave the misreporting region. And this time, even if we are off to the target angle θ_0, we will be closer than before. Therefore, for the interval of possible values of θ' to be at most ϵ, we require at least $\log(\frac{1}{\epsilon})$ rounds of the algorithm, each of which is a constant number of pairwise queries.

□

Proof of Lemma 2. "Under our model, no algorithm can find the maximizer (minimizer) in fewer than $O(\log \frac{1}{\epsilon})$ queries."

For any fixed ϵ, divide the search space θ into bins of length ϵ, resulting in $\lceil \frac{1}{\epsilon} \rceil$ classifiers. When the function evaluated on these classifiers is unimodal, and when the only operation allowed is pairwise comparison, the optimal worst case complexity for finding the argument maximum (of function evaluations) is $O(\log \frac{1}{\epsilon})$ [5], which is achieved by binary search. $\square$

Proposition 5. *Let $(y_1, x_1, h(x_1)), \ldots, (y_n, x_n, h(x_n))$ be n i.i.d. samples from the joint distribution on Y, X, and $h(X)$. Then by Höffding's inequality,*

$$\mathbb{P}\left[\left|\tfrac{1}{n}\sum_{i=1}^n \mathbb{1}[h_i = y_i = 1] - TP(h)\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2}.$$

The same holds for the analogous estimator on TN.

Proof. Direct application of Höffding's inequality. $\square$

Proof of Theorem 2. *"Let $\varphi_{LPM} \ni \phi^* = \mathbf{m}^*$ be the true performance metric. Under Assumption 4, given $\epsilon > 0$, LPM elicitation (Section 5.1) outputs a performance metric $\hat{\phi} = \hat{\mathbf{m}}$, such that $\|\mathbf{m}^* - \hat{\mathbf{m}}\|_\infty < \sqrt{2}\epsilon + \frac{2}{k_0}\sqrt{2k_1\epsilon_\Omega}$."*

We will show this for threshold classifiers, as in the statement of the Assumption 4, but it is not difficult to extend the argument to the case of querying angles. (Involves a good bit of trigonometric identities...)

Recall, the threshold estimator h_δ returns positive if $\eta(x) \geq \delta$, and zero otherwise. Let $\bar{\delta}$ be the threshold which maximizes performance with respect to ϕ, and $C_{\bar{\delta}}$ be its confusion matrix. For simplicity, suppose that $\delta' < \bar{\delta}$. Recall, from Assumption 4 that $\Pr[\eta(X) \in [\bar{\delta} - \frac{k_0}{2k_1}\epsilon, \bar{\delta}]] \leq k_0\epsilon/2$, but $\Pr[\eta(X) \in [\bar{\delta} - \epsilon, \bar{\delta}]] \geq k_0\epsilon$, and therefore

$$\mathbb{P}\Big[\eta(X) \in [\bar{\delta} - \epsilon, \bar{\delta} - \tfrac{k_0}{2k_1}\epsilon]\Big] \geq k_0\epsilon/2$$

Denoting $\phi(C) = \langle \mathbf{m}, C\rangle$, and recalling that $\bar{\delta} = m_{00}/(m_{11} + m_{00})$, expanding the integral, we get

$$\phi(C_{\bar{\delta}}) - \phi(C_{\delta'})$$

$$= \int_{x:\delta' \leq \eta(x) \leq \bar{\delta}} [m_{00}(1 - \eta(x)) - m_{11}\eta(x)]\, \mathrm{d}f_X$$
$$= \int_{x:\bar{\delta} - (\bar{\delta} - \delta') \leq \eta(x) \leq \bar{\delta}} [m_{00}(1 - \eta(x)) - m_{11}\eta(x)]\, \mathrm{d}f_X$$
$$\geq \int_{x:\bar{\delta} - (\bar{\delta} - \delta') \leq \eta(x) \leq \bar{\delta} - \frac{k_0}{2k_1}(\bar{\delta} - \delta')} [m_{00}(1 - \eta(x)) - m_{11}\eta(x)]\, \mathrm{d}f_X$$
$$\geq [(m_{11} + m_{00})\big(\frac{-m_{00}}{m_{00} + m_{11}} + \frac{k_0}{2k_1}(\bar{\delta} - \delta')\big) + m_{00}] \times \int_{x:\bar{\delta} - (\bar{\delta} - \delta') \leq \eta(x) \leq \bar{\delta} - \frac{k_0}{2k_1}(\bar{\delta} - \delta')} \mathrm{d}f_X$$
$$= [(m_{11} + m_{00})\frac{k_0}{2k_1}(\bar{\delta} - \delta')] \times \mathbb{P}[\bar{\delta} - (\bar{\delta} - \delta') \leq \eta(x) \leq \bar{\delta} - \frac{k_0}{2k_1}(\bar{\delta} - \delta')]$$
$$\geq \tfrac{k_0}{2}(\bar{\delta} - \delta') \cdot \tfrac{k_0}{2k_1}(\bar{\delta} - \delta') = \frac{k_0^2}{4k_1}(\bar{\delta} - \delta')^2.$$

Similar results hold when $\delta' > \bar{\delta}$. Therefore, if we have $|\phi(\overline{C}) - \phi(C(\delta'))| < \epsilon_\Omega$, then we must have $|\bar{\delta} - \delta'| < \frac{2}{k_0}\sqrt{k_1\epsilon_\Omega}$. Thus, if we are in a regime where the oracle is mis-reporting the preference ordering, it must be the case that the thresholds are sufficiently close to the optimal threshold.

Again, as in the proof of Theorem 1, when the tolerance ϵ is small, our binary search closes in on a parameter θ' which has $\phi(C_{\delta_{\theta'}})$ within ϵ_Ω of the optimum, but from the above discussion, this also implies that the search interval itself is close to the true value, and thus, the total error in the threshold is at most $\epsilon + \frac{2}{k_0}\sqrt{k_1\epsilon_\Omega}$. Since $\bar{\delta} = m_{00}/(m_{11} + m_{00})$, this bound extends to the cost vector with a factor of $\sqrt{2}$, thus giving the desired result.

We observe that the above theorem actually provide bounds on the slope of the hyperplanes. Thus, the guarantees for LFPM elicitation follow naturally. It only requires that we recover the slope at the upper boundary and lower boundary correctly (within some bounds). This theorem provides those guarantees. Algorithm 3 is independent of oracle queries and thus can be run with high precision, making the solutions of the two systems match. $\square$

Proof of Lemma 3. *"Let h_θ and $\hat{h}_\theta$ be two classifiers estimated using η and $\hat{\eta}$, respectively. Further, let $\bar{\theta}$ be such that $h_{\bar{\theta}} = \arg\max_\theta \phi(h_\theta)$. Then $\|C(\hat{h}_{\bar{\theta}}) - C(h_{\bar{\theta}})\|_\infty = O(\|\hat{\eta}_n - \eta\|_\infty)$."*

Suppose the performance metric of the oracle is characterized by the parameter $\bar{\theta}$. Recall the Bayes optimal classifier would be $h_{\bar{\theta}} = \mathbb{1}[\eta \geq \bar{\delta}]$. Let us assume we are given a classifier $\hat{h}_{\bar{\theta}} = \mathbb{1}[\hat{\eta} \geq \bar{\delta}]$. Notice that the optimal threshold $\bar{\delta}$ is the property of the metric and not the classifier or η. We want to bound the difference in the confusion matrices for these two classifiers. Notice that, by Assumption 3, we can take n sufficiently large so that $\|\eta - \hat{\eta}_n\|_\infty$ is arbitrarily small. Consider the quantity

$$TP(h_{\bar{\theta}}) - TP(\hat{h}_{\bar{\theta}}) = \int_{\eta \geq \bar{\delta}} \eta \, \mathrm{d}f_X - \int_{\hat{\eta} \geq \bar{\delta}} \eta \, \mathrm{d}f_X. \tag{21}$$

Now the maximum loss in the above quantity can occur when, in the region where the classifiers' predictions differ, there $\hat{\eta}$ is less than η with the maximum possible difference. This is equal to

$$\begin{aligned} &\int_{x:\bar{\delta} \leq \eta(x) \leq \bar{\delta} + \|\eta - \hat{\eta}\|_\infty} \eta \, \mathrm{d}f_X \\ &\leq \mathbb{P}[\bar{\delta} \leq \eta(X) \leq \bar{\delta} + \|\eta - \hat{\eta}\|_\infty] \\ &\leq k_1 \|\eta - \hat{\eta}\|_\infty. \qquad \text{(by Assumption 4)} \end{aligned}$$

Similarly, we can look at the maximum gain in the following quantity.

$$TP(\hat{h}_{\bar{\theta}}) - TP(h_{\bar{\theta}}) = \int_{\hat{\eta} \geq \bar{\delta}} \eta \, \mathrm{d}f_X - \int_{\eta \geq \bar{\delta}} \eta \, \mathrm{d}f_X \tag{22}$$

Now the maximum gain in the above quantity can occur when, in the region where the classifiers' predictions differ, there $\hat{\eta}$ is greater than η with the maximum possible difference. This is equal to

$$\begin{aligned} &\int_{x:\bar{\delta} - \|\eta - \hat{\eta}\|_\infty \leq \eta(x) \leq \bar{\delta}} \eta \, \mathrm{d}f_X \\ &\leq \mathbb{P}[\bar{\delta} - \|\eta - \hat{\eta}\|_\infty \leq \eta(X) \leq \bar{\delta}] \\ &\leq k_1 \|\eta - \hat{\eta}\|_\infty. \qquad \text{(by Assumption 4)} \end{aligned}$$

Hence,

$$|TP(\hat{h}_{\bar{\theta}}) - TP(h_{\bar{\theta}})| \leq k_1 \|\eta - \hat{\eta}\|_\infty.$$

Similar arguments apply for TN, which gives us the desired result. $\square$

C Extended Experiments

In this section, we empirically validate the theory and understand the sensitivity due to finite samples.

C.1 Synthetic Data Experiments

We take the same distribution as in (12) with the noise parameter $a = 5$. In the LPM elicitation case, we define a true metric ϕ^* by $\mathbf{m}^* = (m^*_{11}, m^*_{00})$. This defines the query outputs in line 6 of Algorithm 1. Then we run Algorithm 1 to check whether or not we get the same metric. The results for both monotonically increasing and monotonically decreasing LPM are shown in Table 3. We achieve the true metric even for very tight tolerance $\epsilon = 0.02$ radians.

Table 3: Empirical Validation for LPM elicitation at tolerance $\epsilon = 0.02$ radians. ϕ^* and $\hat{\phi}$ denote the true and the elicited metric, respectively.

$\phi^* = \mathbf{m}^*$	$\hat{\phi} = \hat{\mathbf{m}}$	$\phi^* = \mathbf{m}^*$	$\hat{\phi} = \hat{\mathbf{m}}$
(0.98,0.17)	(0.99,0.17)	(-0.94,-0.34)	(-0.94,-0.34)
(0.87,0.50)	(0.87,0.50)	(-0.77,-0.64)	(-0.77,-0.64)
(0.64,0.77)	(0.64,0.77)	(-0.50,-0.87)	(-0.50,-0.87)
(0.34,0.94)	(0.34,0.94)	(-0.17,-0.98)	(-0.17,-0.99)

Next, we elicit LFPM. We define a true metric ϕ^* by $\{(p^*_{11}, p^*_{00}), (q^*_{11}, q^*_{00}, q^*_0)\}$. Then, we run Algorithm 1 with $\epsilon = 0.05$ to find the hyperplane $\bar{\ell}$ and maximizer on ∂C_+, Algorithm 2 with $\epsilon = 0.05$ to find the hyperplane $\underline{\ell}$ and minimizer on ∂C_-, and Algorithm 3 with $n = 2000$ (1000 confusion matrices on both $\partial\mathcal{C}_+$ and $\partial\mathcal{C}_-$ obtained by varying parameter θ uniformly in $[0, \pi/2]$ and $[\pi, 3\pi/2]$) and $\Delta = 0.01$. This gives us the elicited metric $\hat{\phi}$, which we represent by $\{(\hat{p}_{11}, \hat{p}_{00}), (\hat{q}_{11}, \hat{q}_{00}, \hat{q}_0)\}$. In Table 4, we present the elicitation results for LFPMs (column 2). We also present the mean (α) and the standard deviation (σ) of the ratio of the elicited metric $\hat{\phi}$ to the true metric ϕ over the set of confusion matrices (column 3 and 4 of Table 4). As suggested in Corollary 1, if we know the true ratio of p^*_{11}/p^*_{00}, then we can elicit the LFPM up to a constant by only using Algorithm 1 resulting in better estimate of the true metric, because we avoid errors due to Algorithms 2 and 3. Line 1 and line 2 of Table 4 represent F_1 measure and $F_{\frac{1}{2}}$ measure, respectively. In both the cases, we assume the knowledge of $p^*_{11} = 1$. Line 3 to line 6 correspond to some arbitrarily chosen linear fractional metrics to show the efficacy of the proposed method. For a better judgment, we show function evaluations of the true metric and the elicited metric on selected pairs of $(TP, TN) \in \partial\mathcal{C}_+$ (used for Algorithm 3) in Figure 6. The true and the elicited metric are plotted together after sorting values based on slope parameter θ. We see that the elicited metric is a constant multiple of the true metric. The vertical solid line and dashed line corresponds to the *argmax* of the true and the elicited metric, respectively. In Figure 6, we see that the *argmax* of the true and the elicited metrics coincides, thus validating Theorem 1.

C.2 Real-World Data Experiments

In real-world datasets, we do not know $\eta(x)$ and only have finite samples. As a result of these two road blocks, the feasible space $\mathcal{C}$ is not as well behaved as shown in Figure 5, and poses a good challenge for the elicitation task. Now, we validate the elicitation procedure with two real-world datasets.

The datasets are: (a) Breast Cancer (BC) Wiscon-

Table 4: LFPM Elicitation for synthetic distribution (Section C.1) and Magic (M) dataset (Section C.2) with $\epsilon = 0.05$ radians. $(p^*_{11}, p^*_{00}), (q^*_{11}, q^*_{00}, q^*_0)$ denote the true LFPM. $(\hat{p}_{11}, \hat{p}_{00}), (\hat{q}_{11}, \hat{q}_{00}, \hat{q}_0)$ denote the elicited LFPM. α and σ denote the mean and the standard deviation in the ratio of the elicited to the true metric (evaluated on the confusion matrices in $\partial\mathcal{C}_+$ used in Algorithm 3), respectively. We empirically verify that the elicited metric is constant multiple (α) of the true metric.

True Metric	Results on Synthetic Distribution (Section C.1)			Results on Real World Dataset M (Section C.2)		
$(p^*_{11}, p^*_{00}), (q^*_{11}, q^*_{00}, q^*_0)$	$(\hat{p}_{11}, \hat{p}_{00}), (\hat{q}_{11}, \hat{q}_{00}, \hat{q}_0)$	α	σ	$(\hat{p}_{11}, \hat{p}_{00}), (\hat{q}_{11}, \hat{q}_{00}, \hat{q}_0)$	α	σ
(1.00,0.00),(0.50,-0.50,0.50)	(1.00,0.00),(0.25,-0.75,0.75)	0.92	0.03	(1.00,0.00),(0.25,-0.75,0.75)	0.90	0.06
(1.0,0.0),(0.8,-0.8,0.5)	(1.0,0.0),(0.73,-1.09,0.68)	0.94	0.02	(1.0,0.0),(0.72,-1.13, 0.57)	1.06	0.05
(0.8,0.2),(0.3,0.1,0.3)	(0.86,0.14),(-0.13,-0.07, 0.60)	0.90	0.06	(0.23,0.77),(-0.87,0.66,0.76)	0.84	0.09
(0.60,0.40),(0.40,0.20,0.20)	(0.67,0.33),(-0.07,-0.44,76)	0.82	0.05	(0.16,0.84),(-0.89,0.25,0.89)	0.65	0.05
(0.40,0.60),(-0.10,-0.20,0.65)	(0.36,0.64),(-0.21,-0.25,0.73)	0.97	0.01	(0.08,0.92),(-0.75,0.12,0.82)	0.79	0.08
(0.20,0.80),(-0.40,-0.20,0.80)	(0.12, 0.88),(-0.43, 0.002, 0.71)	1.02	0.006	(0.19,0.81),(-0.38,-0.13,0.70)	1.02	0.004

Table 5: LPM elicitation results on real datasets (ϵ in radians). M and BC represent Magic and Breast Cancer dataset, respectively. λ is the regularization parameter in the regularized logistic regression models. The table shows error in terms of the proportion of the number of times when Algorithm 1 (Algorithm 2) failed to recover the true $\mathbf{m}^*(\theta^*)$ within ϵ threshold. The observations made in the main paper are consistent for both the regularized models.

	$\lambda = 10$		$\lambda = 1$	
ϵ	M	BC	M	BC
0.02	0.57	0.79	0.54	0.79
0.05	0.14	0.43	0.36	0.64
0.08	0.07	0.21	0.14	0.57
0.11	0.00	0.07	0.07	0.43

sin Diagnostic dataset [25] containing 569 instances, and (b) Magic (M) dataset [8] containing 19020 instances. For both the datasets, we standardize the attributes and split the data into two parts $\mathcal{S}_1$ and $\mathcal{S}_2$. On $\mathcal{S}_1$, we learn an estimator $\hat{\eta}$ using regularized logistic regression model with regularizing constant $\lambda = 10$ and $\lambda = 1$. We use $\mathcal{S}_2$ for making predictions and computing sample confusion matrices.

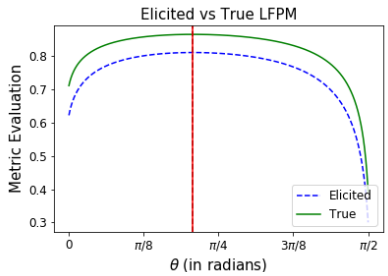
We generated twenty eight different LPMs ϕ^* by generating θ^* (or say, $\mathbf{m}^* = (\cos\theta^*, \sin\theta^*)$). Fourteen from the first quadrant starting from $\pi/18$ radians to $5\pi/12$ radians in step of $\pi/36$ radians. Similarly, fourteen from the third quadrant starting from $19\pi/18$ to $17\pi/12$ in step of $\pi/36$ radians. We then use Algorithm 1 (Algorithm 2) for different tolerance ϵ, for different datasets, and for different regularizing constant λ in order to recover the estimate $\hat{\mathbf{m}}$. We compute the error in terms of the proportion of the number of times when Algorithm 1 (Algorithm 2) failed to recover the true $\mathbf{m}^*$ within ϵ threshold.

We report our results in Table 5. We see improved elicitation for dataset M, suggesting that ME improves with larger datasets. In particular, for dataset M, we elicit all the metrics within threshold $\epsilon = 0.11$ radians. We also observe that $\epsilon = 0.02$ is an overly tight tolerance for both the datasets leading to many failures. This is because the elicitation routine gets stuck at the closest achievable confusion matrix from finite samples, which need not be optimal within the given (small) tolerance. Furthermore, both of these observations are consistent for both the regularized logisitic regression models with regularizer λ.

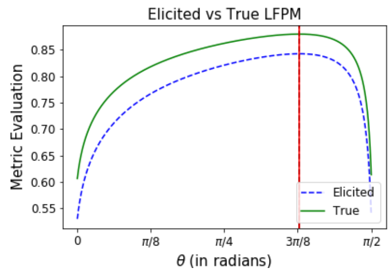
Next, we discuss the case of LFPM elicitation. We use the same true metrics ϕ^* as described in Section C.1 and follow the same process for eliciting LFPM, but this time we work with MAGIC dataset. In Table 4 (columns 5, 6, and 7), we present the elicitation results on MAGIC dataset along with the mean α and the standard deviation σ of the ratio of the elicited metric and the true metric. Again, for a better judgment, we show the function evaluation of the true metric and the elicited metric on the selected pairs of $(TP, TN) \in \partial\mathcal{C}_+$ (used for Algorithm 3) in Figure 7, ordered by the parameter θ. Although we do observe that the *argmax* is different in two out of six cases (see Sub-figure (b) and Sub-figure (c)) due to finite samples, elicited LFPMs are almost equivalent to the true metric up to a constant.

D Monotonically Decreasing Case

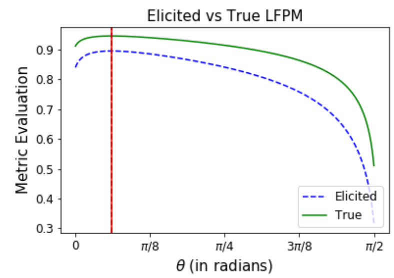
Even if the oracle's metric is monotonically decreasing in TP and TN, we can figure out the supporting hyperplanes at the maximizer and the minimizer. It would require to pose one query $\Omega(C^*_{\pi/4}, C^*_{5\pi/4})$. The response from this query determines whether we want to search over $\partial\mathcal{C}_+$ or $\partial\mathcal{C}_-$ and apply Algorithms 1 and 2 accordingly. In fact, if $C^*_{\pi/4} \prec C^*_{5\pi/4}$, then the metric is monotonically decreasing, and we search for the maximizer on the lower boundary $\partial\mathcal{C}_-$. Similarly if the converse holds, then we search over $\partial\mathcal{C}_+$ as discussed in the main paper.



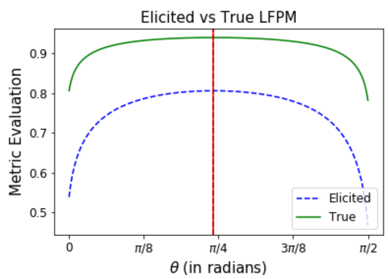
(a) Table 4, Line 1, Column 2



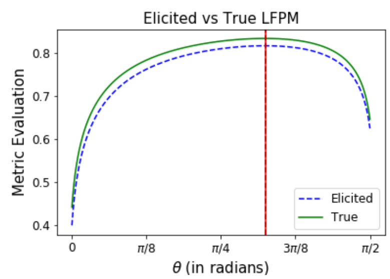
(b) Table 4, Line 2, Column 2



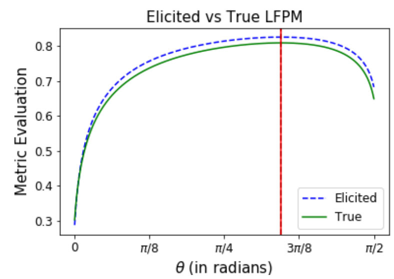
(c) Table 4, Line 3, Column 2



(d) Table 4, Line 4, Column 2

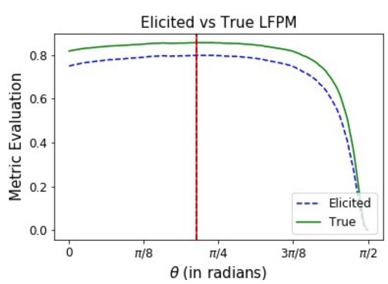


(e) Table 4, Line 5, Column 2

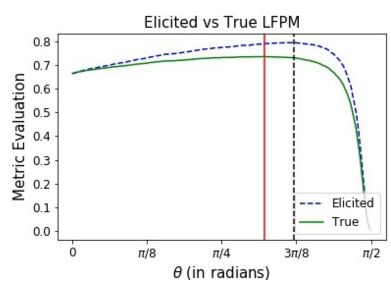


(f) Table 4, Line 6, Column 2

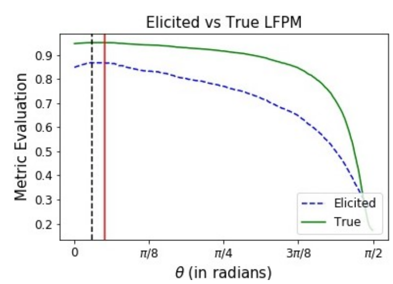
Fig. 6: True and elicited LFPMs for synthetic distribution from Table 4. The solid green curve and the dashed blue curve are the true and the elicited metric, respectively. The solid red and the dashed black vertical lines represent the maximizer of the true metric and the elicited metric, respectively. We see that the elicited LFPMs are constant multiple of the true metrics with the same maximizer (solid red and dashed black vertical lines overlap in all the cases).



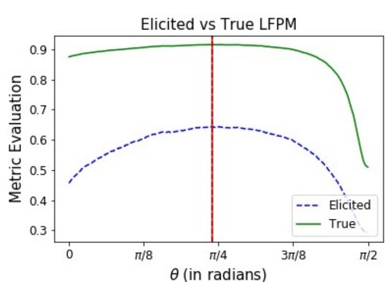
(a) Table 4, Line 1, Column 5



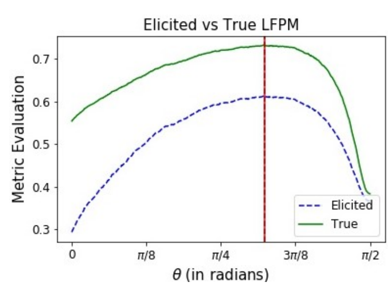
(b) Table 4, Line 2, Column 5



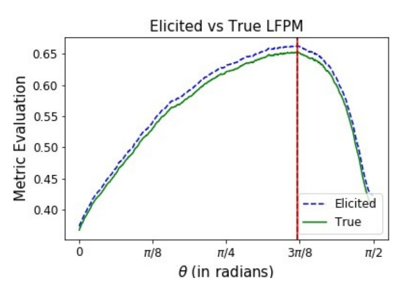
(c) Table 4, Line 3, Column 5



(d) Table 4, Line 4, Column 5



(e) Table 4, Line 5, Column 5



(f) Table 4, Line 6, Column 5

Fig. 7: True and elicited LFPMs for dataset M from Table 4. The solid green curve and the dashed blue curve are the true and the elicited metric, respectively. The solid red and the dashed black vertical lines represent the maximizer of the true metric and the elicited metric, respectively. We see that the elicited LFPMs are constant multiple of the true metrics with almost the same maximizer (solid red and dashed black vertical lines overlap except for two cases).