A Proof of Lemma 1

We first introduce the following Lemma, which shows that w_1^t and w_2^t can be written in closed-forms in terms of $(w_1^0, w_2^0, M_{\rm tr}, \alpha, t)$:

Lemma 7. *Let $M_{\rm tr} = U\Sigma V^T$ be (any of) its SVD such that $U \in \mathbb{R}^{d\times k}, \Sigma \in \mathbb{R}^{k\times k}, V \in \mathbb{R}^{k\times k}$, $U^TU = V^TV = VV^T = I$. Then for any $t \ge 0$*

$$\begin{aligned} w_1^t &= \frac{1}{2}V\left(\Lambda^{+,t}V^Tw_1^0 + \Lambda^{-,t}U^Tw_2^0\right) , \\ w_2^t &= \frac{1}{2}U\left(\Lambda^{-,t}V^Tw_1^0 + \Lambda^{+,t}U^Tw_2^0\right) - UU^Tw_2^0 + w_2^0 . \end{aligned} \tag{4}$$

where we define $\Lambda^{+,t} = (I+\alpha\Sigma)^t + (I-\alpha\Sigma)^t$ and $\Lambda^{-,t} = (I+\alpha\Sigma)^t - (I-\alpha\Sigma)^t$.

Proof. We start with stating the following facts for $\Lambda^{+,t}$ and $\Lambda^{-,t}$:

$\Lambda^{+,0} = 2I$, $\Lambda^{-,0} = 0$ and for any $t \ge 0$

$$\begin{aligned} \Lambda^{+,t+1} &= \Lambda^{+,t} + \alpha\Sigma\Lambda^{-,t} , \\ \Lambda^{-,t+1} &= \Lambda^{-,t} + \alpha\Sigma\Lambda^{+,t} . \end{aligned}$$

Now we prove (4) by induction. When $t = 0$, $w_1^0 = VV^Tw_1^0$ and $w_2^0 = UU^Tw_2^0 - UU^Tw_2^0 + w_2^0$ so (4) holds for $t = 0$. Assume Lemma (4) holds for t then consider the next step $t+1$:

$$\begin{aligned} w_1^{t+1} &= w_1^t + \alpha M_{\rm tr}^T w_2^t \\ &= \frac{1}{2}V\left(\Lambda^{+,t}V^Tw_1^0 + \Lambda^{-,t}U^Tw_2^0\right) \\ &+ \alpha V\Sigma U^T\left(\frac{1}{2}U\left(\Lambda^{-,t}V^Tw_1^0 + \Lambda^{+,t}U^Tw_2^0\right)\right. \\ &\left. - UU^Tw_2^0 + w_2^0\right) \\ &= \frac{1}{2}V\left(\Lambda^{+,t}V^Tw_1^0 + \Lambda^{-,t}U^Tw_2^0\right. \\ &\left. + \alpha\Sigma\Lambda^{-,t}V^Tw_1^0 + \alpha\Sigma\Lambda^{+,t}U^Tw_2^0\right) \\ &= \frac{1}{2}V\left(\Lambda^{+,t+1}V^Tw_1^0 + \Lambda^{-,t+1}U^Tw_2^0\right) . \end{aligned}$$

Similarly, we can show

$$\begin{aligned} w_2^{t+1} &= w_2^t + \alpha M_{\rm tr} w_1^t \\ &= \frac{1}{2}U\left(\Lambda^{-,t+1}V^Tw_1^0 + \Lambda^{+,t+1}U^Tw_2^0\right) \\ &- UU^Tw_2^0 + w_2^0 . \end{aligned}$$

Thus (4) holds for all $t \ge 0$. □

Proof of Lemma 1. Taking $w_2^0 = 0$ in Lemma 7 we can write $w_1^t = \frac{1}{2}V\Lambda^{+,t}V^Tw_1^0$ and $w_2^t = \frac{1}{2}U\Lambda^{-,t}V^Tw_1^0$

For $1 \le i \le m$, $\sigma_i = \sigma_1$ thus

$$\lim_{t\to+\infty} \frac{(1+\alpha\sigma_i)^t}{(1+\alpha\sigma_1)^t} = 1 . \tag{5}$$

For $m < i \le k$, $\sigma_i < \sigma_1$ thus

$$\lim_{t\to+\infty} \frac{(1+\alpha\sigma_i)^t}{(1+\alpha\sigma_1)^t} = 0 . \tag{6}$$

For any $1 \le i \le k$, we have $\frac{1-\alpha\sigma_i}{1+\alpha\sigma_1} \le \frac{1}{1+\alpha\sigma_1} < 1$ and $\frac{1-\alpha\sigma_i}{1+\alpha\sigma_1} \ge \frac{1-\alpha\sigma_1}{1+\alpha\sigma_1} = -1 + \frac{2}{1+\alpha\sigma_1} > -1$ thus

$$\lim_{t\to+\infty} \frac{(1-\alpha\sigma_i)^t}{(1+\alpha\sigma_1)^t} = 0 . \tag{7}$$

Applying (5)—(7) to compute the limits in (2) gives the result in Lemma 1. □

B Proof of Theorem 2

Proof. For any vector $z \in \mathbb{R}^m$ such that $\|z\|_2 = 1$, we have

$$\begin{aligned} M_{\rm tr}V_{:m}z &= U\Sigma V^TV_{:m}z = \sigma_1 U_{:m}z , \\ M_{\rm tr}^TU_{:m}z &= V\Sigma U^TU_{:m}z = \sigma_1 V_{:m}z , \end{aligned}$$

Since

$$\begin{aligned} \|V_{:m}z\|_2^2 &= z^TV_{:m}^TV_{:m}z = 1 , \\ \|U_{:m}z\|_2^2 &= z^TU_{:m}^TU_{:m}z = 1 \end{aligned}$$

we know that $(U_{:m}z, V_{:m}z)$ is also a pair of left-right singular vectors with singular value σ_1. Therefore, when $V_{:m}^Tw_1^0 \in \mathbb{R}^m$ is non-zero $\left(\frac{U_{:m}V_{:m}^Tw_1^0}{\|V_{:m}^Tw_1^0\|_2}, \frac{V_{:m}V_{:m}^Tw_1^0}{\|V_{:m}^Tw_1^0\|_2}\right)$ is also such a pair. Following (3) we have

$$\begin{aligned} \frac{\mathcal{F}^\infty(x,y,w_1^0,D_{\rm tr})}{\|V_{:m}^Tw_1^0\|_2^2} &= \left(\frac{V_{:m}V_{:m}^Tw_1^0}{\|V_{:m}^Tw_1^0\|_2}\right)^T M_{x,y}^T \left(\frac{U_{:m}V_{:m}^Tw_1^0}{\|V_{:m}^Tw_1^0\|_2}\right) \\ &\ge \min_{(u,v)\in UV_1^{M_{\rm tr}}} v^TM_{x,y}^Tu \end{aligned} \tag{8}$$

for any w_1^0 such that $V_{:m}^Tw_1^0 \ne \vec{0}$.

When $w_1^0 \sim \mathcal{N}(0, b^2I_k)$, for any fixed $V_{:m}$ satisfying $V_{:m}^TV_{:m} = I_m$, the random variable $V_{:m}^Tw_1^0$ also follows a normal distribution:

$$\mathbb{E}\left[V_{:m}^Tw_1^0(V_{:m}^Tw_1^0)^T\right] = V_{:m}^T\mathbb{E}\left[w_1^0{w_1^0}^T\right]V_{:m} = b^2I_m$$

hence $V_{:m}^Tw_1^0 \sim \mathcal{N}(0, b^2I_m)$.

Applying the fact that $\bar{\mathcal{E}}(\cdot) \le 1$ is non-increasing and $\bar{\mathcal{E}}(\alpha x) = \bar{\mathcal{E}}(x)$ for any $\alpha > 0$ we can upper bound (3) by

$\mathcal{E}^\infty_{\rm Convk}(\mathcal{D})$

$$
\begin{aligned}
&= \mathbb{E}_{w_1^0, D_{\mathrm{tr}},(x,y)}\left[\bar{\mathcal{E}}\left(\mathcal{F}^\infty(x,y,w_1^0,D_{\mathrm{tr}})\right)\right] \\
&= \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\mathbb{E}_{w_1^0}\left[\bar{\mathcal{E}}\left(\mathcal{F}^\infty(x,y,w_1^0,D_{\mathrm{tr}})\right)\right]\right] \\
&= \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\Pr\left(V_{:m}^T w_1^0 = \vec{0}\right)\mathbb{E}_{w_1^0}\left[\bar{\mathcal{E}}\left(\mathcal{F}^\infty\right)\left|V_{:m}^T w_1^0 = \vec{0}\right]\right.\right. \\
&+ \Pr\left(V_{:m}^T w_1^0 \neq \vec{0}\right)\mathbb{E}_{w_1^0}\left.\left.\left[\bar{\mathcal{E}}\left(\mathcal{F}^\infty\right)\left|V_{:m}^T w_1^0 \neq \vec{0}\right]\right]\right. \\
&= \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\mathbb{E}_{w_1^0}\left[\bar{\mathcal{E}}\left(\frac{\mathcal{F}^\infty}{\|V_{:m}^T w_1^0\|_2^2}\right)\left|V_{:m}^T w_1^0 \neq \vec{0}\right]\right]\right. \\
&\leq \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\bar{\mathcal{E}}\left(\min_{(u,v)\in UV_1^{M_{\mathrm{tr}}}} v^T M_{x,y}^T u\right)\right].
\end{aligned}
$$

□

C Perron-Frobenius Theorem

Let $A \in \mathbb{R}^{k\times k}$ be a non-negative square matrix[7]:

- **Definition:** A is *primitive* if there exists a positive integer t such that $A_{ij}^t > 0$ for all i, j.
- **Definition:** A is *irreducible* if for any i, j there exists a positive integer t such that $A_{ij}^t > 0$.
- **Definition:** Its *associated graph* $\mathcal{G}_A = (V, E)$ is defined to be a directed graph with $V = \{1, ..., k\}$ and $(i, j) \in E$ iff $A_{ij} \neq 0$. $\mathcal{G}_A$ is said to be *strongly connected* if for any i, j there is path from i to j.
- **Property:** A is irreducible iff $\mathcal{G}_A$ is strongly connected.
- **Property:** If A is irreducible and has at least one non-zero diagonal element then A is primitive.
- **Property:** If A is primitive then its first eigenvalue is unique ($\lambda_1 > \lambda_2$) and the corresponding eigenvector is all-positive (or all-negative up to sign flipping).

D Proof of Theorem 4

Proof. Following (3) and let

$$\hat{\mathcal{E}}(x,y,D_{\mathrm{tr}}) = \bar{\mathcal{E}}\left(\min_{(u,v)\in UV_1^{M_{\mathrm{tr}}}} v^T M_{x,y}^T u\right) \leq 1$$

we have

$$
\begin{aligned}
&\mathcal{E}_{\mathrm{Convk}}^\infty(\mathcal{D}) \leq \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\hat{\mathcal{E}}(x,y,D_{\mathrm{tr}})\right] \\
&= \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\left(\mathbb{I}\left\{\Omega^c(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\} + \mathbb{I}\left\{\Omega(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\right)\right. \\
&\left.\hat{\mathcal{E}}(x,y,D_{\mathrm{tr}})\right] \\
&\leq \mathbb{E}_{D_{\mathrm{tr}}}\left[\mathbb{I}\left\{\Omega^c(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\right] \\
&+ \mathbb{E}_{D_{\mathrm{tr}},(x,y)}\left[\mathbb{I}\left\{\Omega(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\hat{\mathcal{E}}(x,y,D_{\mathrm{tr}})\right] \\
&= \Pr\left(\Omega^c(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right) \\
&+ \mathbb{E}_{D_{\mathrm{tr}},l\sim\mathcal{U}[d]}\left[\mathbb{I}\left\{\Omega(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\hat{\mathcal{E}}(e_l,1,D_{\mathrm{tr}})\right] \qquad (9)
\end{aligned}
$$

Now look at the second term in (9). If $M_{\mathrm{tr}}^T M_{\mathrm{tr}}$ is primitive then its first eigenvalue $\lambda_1 = \sigma_1^2$ is unique ($\sigma_1 > \sigma_2$) and the corresponding eigenvector v is all positive (or all negative if we flip the sign of v and u, which does not change the sign of $v^T M^T u$ thus it is safe to assume $v > 0$). $u = M_{\mathrm{tr}} v/\sigma_1$ gives that u is also unique and non-negative. Since $M_{x,y}$ is also non-negative we have $v^T M_{x,y}^T u \geq 0$ for any x, y. Therefore,

$$
\begin{aligned}
\hat{\mathcal{E}}(x,y,D_{\mathrm{tr}}) &= \bar{\mathcal{E}}\left(v^T M_{x,y}^T u\right) \\
&= \mathbb{I}\left\{v^T M_{x,y}^T u < 0\right\} + \frac{1}{2}\mathbb{I}\left\{v^T M_{x,y}^T u = 0\right\} \\
&= \frac{1}{2}\mathbb{I}\left\{v^T M_{x,y}^T u = 0\right\}.
\end{aligned}
$$

From $u = M_{\mathrm{tr}} v/\sigma_1$ and $v > 0$ we know that $u_i > 0$ iff there exists $1 \leq j \leq k$ such that $(M_{\mathrm{tr}})_{i,j} > 0$, which is equivalent to that there exists $i \leq l < i + k$ such that $l \in S_{\mathrm{tr}}$. Also for $x = e_l$ ($y = 1$), according to the definition of $M_{x,y}$ and the fact that $v > 0$ we have $v^T M_{e_l,1}^T u > 0$ iff there exists $l - k < i \leq l$ such that $u_i > 0$. So we have

$$v^T M_{e_l,1}^T u > 0 \iff \exists l' \in \bigcup_{l-k<i\leq l} [i, i+k) \text{ s.t. } l' \in S_{\mathrm{tr}}$$

Since $v^T M_{e_l,1}^T u \geq 0$ and $\bigcup_{l-k<i\leq l}[i, i+k) = (l-k, l+k)$ we have

$$v^T M_{e_l,1}^T u = 0 \iff \forall l' \in S_{\mathrm{tr}}, |l' - l| \geq k\,.$$

Now we have proved that, if $M_{\mathrm{tr}}^T M_{\mathrm{tr}}$ is primitive then

$$\hat{\mathcal{E}}(e_l,1,D_{\mathrm{tr}}) = \frac{1}{2}\mathbb{I}\left\{\forall l' \in S_{\mathrm{tr}}, |l'-l| \geq k\right\},$$

which means that

$$\mathbb{I}\left\{\Omega(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\hat{\mathcal{E}}(e_l,1,D_{\mathrm{tr}}) \leq \frac{1}{2}\mathbb{I}\left\{\forall l' \in S_{\mathrm{tr}}, |l'-l| \geq k\right\}$$

holds for any D_{tr}. Therefore

$$
\begin{aligned}
&\mathbb{E}_{D_{\mathrm{tr}},l\sim\mathcal{U}[d]}\left[\mathbb{I}\left\{\Omega(M_{\mathrm{tr}}^T M_{\mathrm{tr}})\right\}\hat{\mathcal{E}}(e_l,1,D_{\mathrm{tr}})\right] \\
&\leq \frac{1}{2}\mathbb{E}_{D_{\mathrm{tr}},l\sim\mathcal{U}[d]}\left[\mathbb{I}\left\{\forall l' \in S_{\mathrm{tr}}, |l'-l| \geq k\right\}\right] \\
&= \frac{1}{2}\mathbb{E}_{l\sim\mathcal{U}[d]}\left[\Pr\left(\forall l' \in S_{\mathrm{tr}}, |l'-l| \geq k\right)\right]
\end{aligned}
$$

which concludes the proof. □

E Proof of Lemma 5

Proof. If $k \leq i \leq d$ and $i - 1, i \in S_{\mathrm{tr}}$ then for any $1 \leq j \leq k$ we have $(M_{\mathrm{tr}})_{i-j,j} > 0$ and $(M_{\mathrm{tr}})_{i-j+1,j} > 0$,

[7] `https://en.wikipedia.org/wiki/Perron-Frobenius_theorem` .

which also means that for any $1 \le j < k$ we have $(M_{\text{tr}})_{i-j,j} > 0$ and $(M_{\text{tr}})_{i-j,j+1} > 0$. Since every two adjacent columns have at least one common non-zero position what we have is $(M_{\text{tr}}^T M_{\text{tr}})_{j,j+1} > 0$ and $(M_{\text{tr}}^T M_{\text{tr}})_{j+1,j} > 0$ for all $1 \le j < k$. So its associated graph $\mathcal{G}_{M_{\text{tr}}^T M_{\text{tr}}}$ is strongly connected thus $M_{\text{tr}}^T M_{\text{tr}}$ is irreducible. It is also true that all diagonal elements of $M_{\text{tr}}^T M_{\text{tr}}$ are positive since every column of M_{tr} must contain at least one non-zero element. Now we have proved that $M_{\text{tr}}^T M_{\text{tr}}$ is primitive because it is irreducible and has at least one non-zero element on its diagonal. □

F Proof of Proposition 6

Proof. Let $n = |S_{\text{tr}}|$. Then given the conditions in this proposition we can see that any column in M_{tr} has exactly n non-zero entries with value $1/n$ and any two columns in M_{tr} has no overlapping non-zero positions. Hence we have $M_{\text{tr}}^T M_{\text{tr}} = \frac{1}{n} I_k$ so that $m = k$ in Lemma 1 and $VV^T = I$. Applying Lemma 1 we have $w_1^\infty = w_1^0$ and $w_2^\infty = n M_{\text{tr}} w_1^0$. Then for any $x = e_l$ we have

$$y f_{w^\infty}(x) = {w_1^\infty}^T M_{x,y}^T w_2^\infty = n {w_1^0}^T A_x^T M_{\text{tr}} w_1^0 .$$

For x to be correctly classified we need $y f_{w^\infty}(x) > 0$. We will show that this is guaranteed only when $l \in S_{\text{tr}}$, i.e. x or $-x \in D_{\text{tr}}$.

Since for any $l, l' \in S_{\text{tr}}$, $|l - l'| \ge 2k$ we know that there exist at most one $l' \in S_{\text{tr}}$ such that $|l - l'| < k$.

If there does not exist such l' then $A_x^T M_{\text{tr}} = 0$ and $y f_{w^\infty}(x) = 0$, which means x is classified randomly.

If there exists a unique l' such that $|l - l'| < k$ and let $s = |l - l'|$, we have that

$$y f_{w^\infty}(x) = n {w_1^0}^T A_x^T M_{\text{tr}} w_1^0 = \sum_{i=1}^{k-s} w_{1,i}^0 w_{1,i+s}^0 .$$

When $l \in S_{\text{tr}}$, which means $s = 0$, we have $y f_{w^\infty}(x) = {w_1^0}^T w_1^0 > 0$ when $w_1^0 \ne 0$ (which holds almost surely).

When $0 < s < k$ it is not guaranteed that $\sum_{i=1}^{k-s} w_{1,i}^0 w_{1,i+s}^0 > 0$ under $w_1^0 \sim \mathcal{N}(0, b^2 I)$. Actually we can show that the distribution of this quantity is symmetric around 0: For any s we can draw a graph with k nodes and every $(i, i+s)$ forms an edge. This graph contains s independent chains so we can choose a set of nodes $S \subset [k]$ such that for any edge exactly one of the two nodes is contained in S. Now for any w_1^0 if we flip the sign at the positions that belong to S then the sign of $\sum_{i=1}^{k-s} w_{1,i}^0 w_{1,i+s}^0$ is also flipped. With $w_1^0 \sim \mathcal{N}(0, b^2 I)$ this indicates that $P(\sum_{i=1}^{k-s} w_{1,i}^0 w_{1,i+s}^0 > 0) = 1/2$.

Now we have shown that, under the condition in this proposition, a data sample is correctly classified by Conv-k with w^∞ if and only if this sample appears in the training set. Otherwise it has only a half change to be correctly classified. This generalization behavior is exactly the same as Model-1-Layer in Task-Cls, which concludes the proof.

□

G A Supporting Evidence for Interpreting Conv-Filters as a Data Adaptive Bias

We have shown that, different from typical regularizations, the bias itself may require some samples to be built up (see Figure 4(b)). We conjecture that convolution layer adds a data adaptive bias: The set of possible filters forms a set of biases. With a few number of samples gradient descent is able to figure out which bias(filter) is more suitable for the dataset. Then the identified bias can play as a prior knowledge to reduce the sample complexity. We provide another evidence for this: Let the dataset contains all $e_l, l \in [d]$ while $y_{e_l} = +1$ if l is odd and -1 is l is even. Model-Conv-k is still able to outperform Model-1-Layer on this task (see Figure 7). We observe that the sign of the learned filter looks like (+, -, +, -, ...) in contrast to the ones learned in our three tasks, which are likely to be all positive or all negative. This indicates that, besides spatial shifting invariance, jointly training the convolutional filter can exploit a broader set of structures and be adaptive to different data distributions.

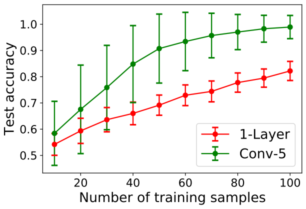
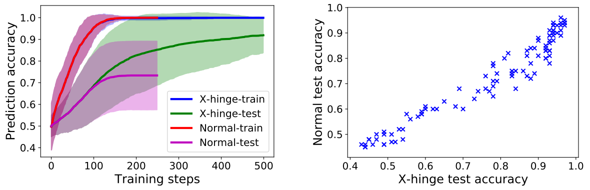


Figure 7: Classifying even v.s. odd non-zero position. Settings are the same as in Figure 2.

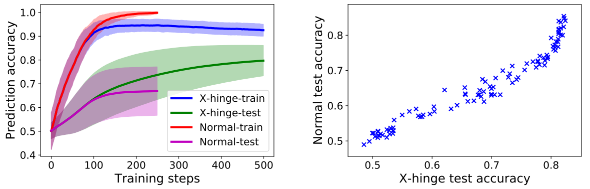
H Correlation Between Normal-hinge and X-hinge under Different Initializations

Figure 8 and 9 shows the variance introduced by weight initialization is also strongly correlated under two losses in Task-1stCtrl and Task-3rdCtrl. Figure 9(a) looks a bit different from the other two tasks because the extreme hinge loss is biased and w^∞ may not able to separate the training samples in Task-3rdCtrl. But the strong correlation between the normal hinge loss and the extreme hinge loss under different weight initializations still holds.



(a) Convergence with t. (b) Correlation at $t = 150$.

Figure 8: The effect of weight initialization in Task-1stCtrl. We fix $d = 100$, $n = 30$ and train Model-Conv-k with 100 different random initializations using both losses.



(a) Convergence with t. (b) Correlation at $t = 150$.

Figure 9: The effect of weight initialization in Task-3stCtrl. We fix $d = 100$, $n = 50$ and train Model-Conv-k with 100 different random initializations using both losses.

I The bias of X-Hinge in Task-3rdCtrl and Potential Practical Indications

In Figure 9(a) we observe that running gradient descent may not be able to achieve 0 training error even if the samples are linearly separable. To explain this, simply consider a training set with 3 samples and $k = 1, d = 4$: $x_1 = [-1, 1, 0, 0], x_2 = [0, -1, 1, 0], x_3 = [0, 0, -1, 1]$. All labels are positive. Then $M_{\text{tr}} = [-1/3, 0, 0, 1/3]$. If we optimize the X-hinge loss then the network has no intent to classify x_2 correctly.

Notice that in Figure 9(a), under X-hinge, the generalization performance is still improving even after the training accuracy starts to decrease. We conjecture that this indicates a new way of interpreting the role of regularization in deep nets. On real datasets we typically use sigmoid with cross entropy loss which can be viewed and a smoothed version of the hinge loss. We say a data sample is *active* during training if $yf(x)$ is small so that the gradient for fitting (x, y) is salient since it is not well fit yet. With X-hinge all samples are "equality active". One message delivered by our observation is that having more samples to be "active" during training will make convolution filters have better generalization property, but may hurt with training data fitting. In practice we cannot recommend using X-hinge loss since the network will fail to fit the training set if we keep *all* samples to be equally "active". But we can view this as a trade off when using logistic loss: keeping more samples to be "active" during training with gradient descent will help with some generalization property (e.g. better Conv filters) but cause underfitting. For regularization we may want to keep as many samples to be active as possible while still be able to fit the training samples. This provides a new view of the role of regularization: Taking weight norm regularization as an example, traditional interpretation is that controlling the weight norm will reduce the capacity of neural nets, which may not be sufficient to explain non-overfitting in very large nets. The new potential interpretation is that, if we keep the weight norm to be small during training, the training samples are more "active" during gradient descent so that better convolution filters can be learned for generalization purposes. Verifying this conjecture on real datasets will be an interesting future direction.