$\boxed{a @ b}$

$\boxed{\text{np.tensordot}(\quad)} \iff (a.T @ b).T$

$\boxed{(Ax)'_x = A^T} = \Sigma$

np.dot( )

SGD

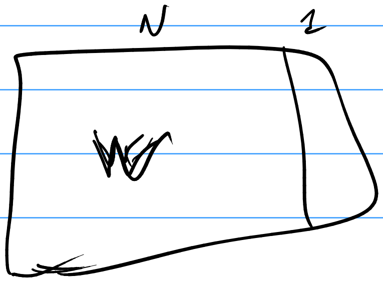
```
for p in params:
    p.data -= τ p.grad.data
p.data = p.data - ...
```

Model → [self.w1, self.bias1, self.w2, ..]

$28^2 + 1$

$$x_1 w_1 + \dots + x_N w_N + 1 \cdot w_{N+1} = x \cdot w$$

$x = (\dots, 1)$

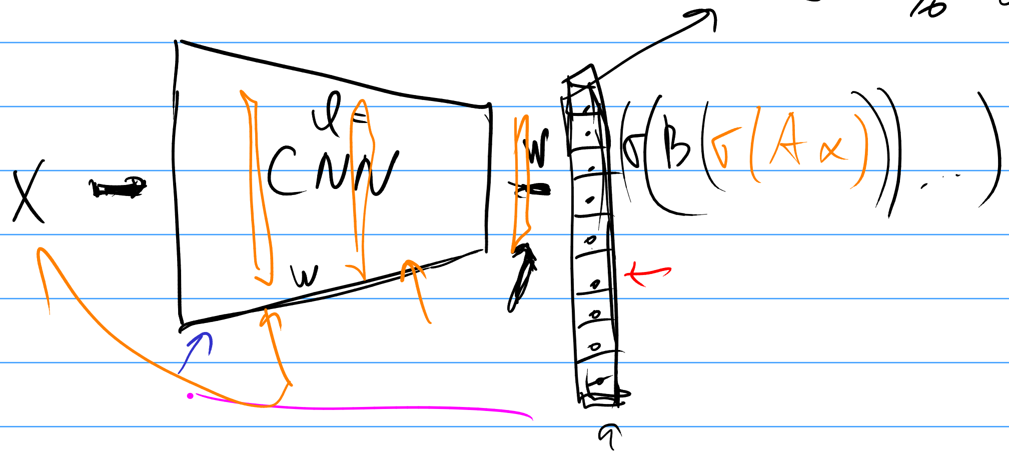


→ random

→ W

Transfer learning

ImageNet   1000 × 1000 = 1M   80 % accuracy



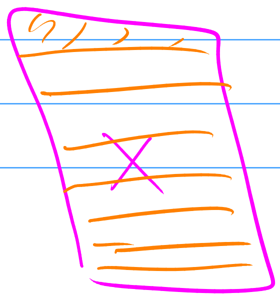
$$\text{logits} = \overbrace{\varphi(X)}^{2048} \cdot \overset{2048 \times 1000}{W} \approx$$

5 классов

$$X \rightarrow \boxed{\varphi(X)\ \overset{2048 \times 5}{W}}$$

random

$\boxed{DT}$, $\boxed{RF}$, $\boxed{GB}$



Finetuning

$$\underbrace{\varphi(X)}_{\uparrow}\cdot \underset{\uparrow}{W}$$

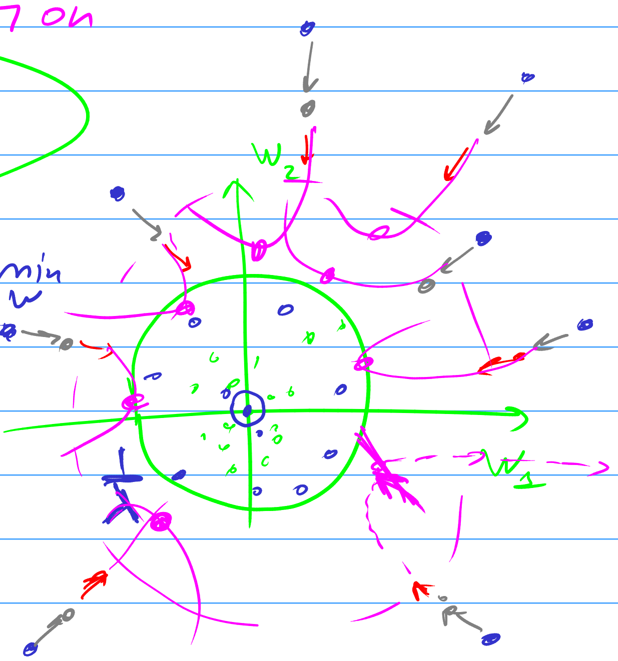
$lr=10^{-4}$ $lr=10^{-3}$

L2 - regularization

$$f_w(x) = \underbrace{XW}$$

$$LOSS = \mathcal{L}(XW, Y) + \lambda\|W\|_2^2 \to \min_w$$

$$\|W\|_2^2 = \sum_{ij} w_{ij}^2 \to \min$$



$$\frac{\partial}{\partial x}\Big\{\ 10^6\cdot X$$

$$\boxed{f - \text{лин.}}$$

L_2-regulanizatin ⟺ early stopping

(best ep - ep) ≥ es.

$$\frac{\partial\|W\|^2}{\partial W} = \dots$$

$$B\overset{D}{\boxed{Y}}\quad D\overset{F}{\boxed{W}}$$

$$\frac{\partial}{\partial W}\left(\underbrace{XW}\right) \qquad \frac{\partial}{\partial w_{lm}}\left(\sum_k x_{ik}w_{kj}\right) =$$

$$\boxed{\frac{\partial x}{\partial x}, \frac{\partial c}{\partial x}}$$

$$\delta_{ij} = \begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}$$

$$= \sum_k x_{ik}\underbrace{\frac{\partial w_{kj}}{\partial w_{lm}}} = \boxed{\sum_k x_{ik}\,\delta_{kl}\,\delta_{jm}}$$

$$\left(\frac{\partial (XW)_{ij}}{\partial w_{lm}}\right).$$