

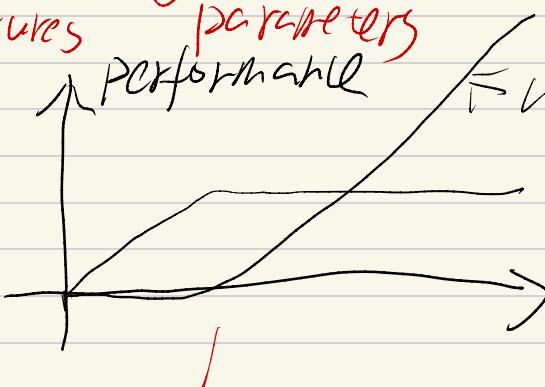
Feb/4/2025

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^n \ell(y_i, f(x_i, \theta))$$

$$f(x, \theta) \in DNN$$

features parameters

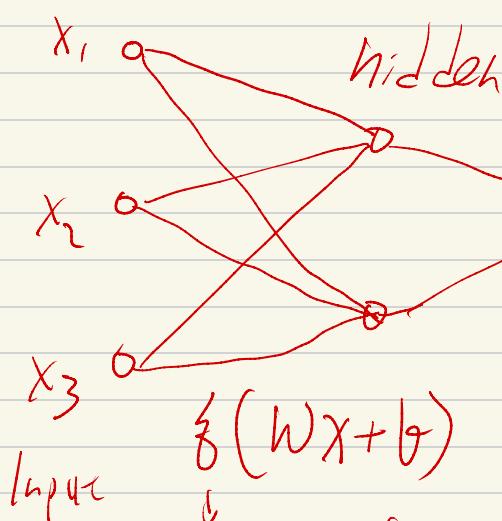
(0)



where we are (2018)

Data/Complete

(1) what is $f(\cdot)$?



$$f(x, \theta) = \delta_2(w_2 \delta_1(w_1 x + b_1) + b_2)$$
$$\Pr(Y=1|X)$$

$\delta(wx + b)$
activation function
~~ReLU = max(x, 0)~~

② Why is the loss $\mathcal{L}(\hat{y}, y)$

Regression: $(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$

Classification: $(\hat{y}, y) = -y \log \hat{y} - (1-y) \log(1-\hat{y})$

where $\hat{y} = f(x, \theta)$

③ How do we find $\hat{\theta}$ $\mathcal{L}(\theta)$

Gradient Descent: $\hat{\theta}_{t+1} = \hat{\theta}_t - \alpha \nabla_{\theta} \mathcal{L}(\hat{\theta}_t)$

④ How do we find $\nabla_{\theta} \mathcal{L}(\theta)$

(i) $y = f(z)$ $z = g(x)$

$$\frac{dy}{dx} = \frac{df}{dz} \Big|_{z=g(x)} \cdot \frac{dg}{dx} \Big|_x$$

(ii) SGD $\{\mathbf{1}, \dots, \mathbf{B}\} \subseteq \{\mathbf{1}, \dots, n\}$

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \alpha \frac{1}{B} \sum_{i=1}^B \nabla_{\theta} \mathcal{L}(x_i; f(x_i; \hat{\theta}_t))$$

Momentum

$$\vec{m}_t \rightarrow \vec{m}_{t+1} = \mu \vec{m}_t + (1-\mu) \vec{g}_t$$

Adam = Momentum + Adaptive learning rate

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \alpha \vec{m}_t$$

(iii) Optimization Tricks 

{ a) Skip connection / ResNet.
 g_1, g_2, \dots, g_k (gradient skipping)

$$\nabla_{\theta} f_i(\theta) = g_i \quad f_i = \theta + h_i(\theta)$$

$$\nabla_{\theta} f_i = \text{Id} + \nabla_{\theta} h_i \quad \text{residual}$$

(b) Initialization: θ

$$(w, b) \sim N(0, \frac{1}{n_L})$$

(c) Batch Normalization:

$$\frac{x_i - \bar{x}_i}{\hat{s}_0(x_i)}$$

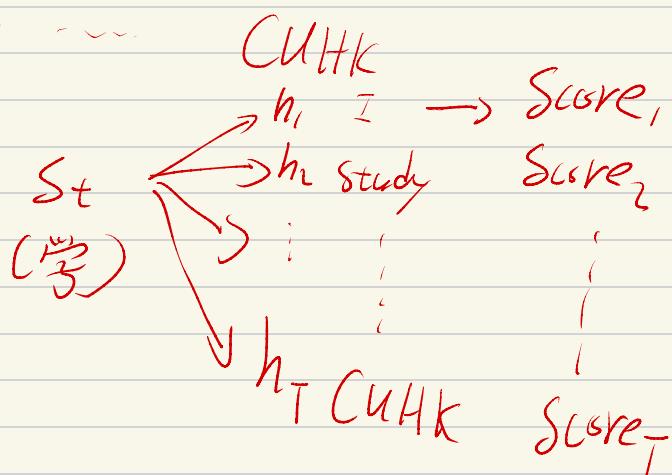
w: $d_{in} \times d_{out}$

b: d_{out}

$h_1 \rightarrow h_2 \rightarrow \dots \rightarrow h_T$ encoder

I Study

Decoder



$$a_t = \sum_{i=1}^T \text{Score}_i \cdot h_i$$

$$\begin{aligned} \text{Score}_i &= \text{Softmax}(s_t \cdot h_i^\top) \\ &= \frac{\exp(s_t \cdot h_i^\top)}{\sum_{j=1}^T \exp(s_t \cdot h_j^\top)} \end{aligned}$$

