Video 16

Wednesday, October 21, 2020 8:53 PM

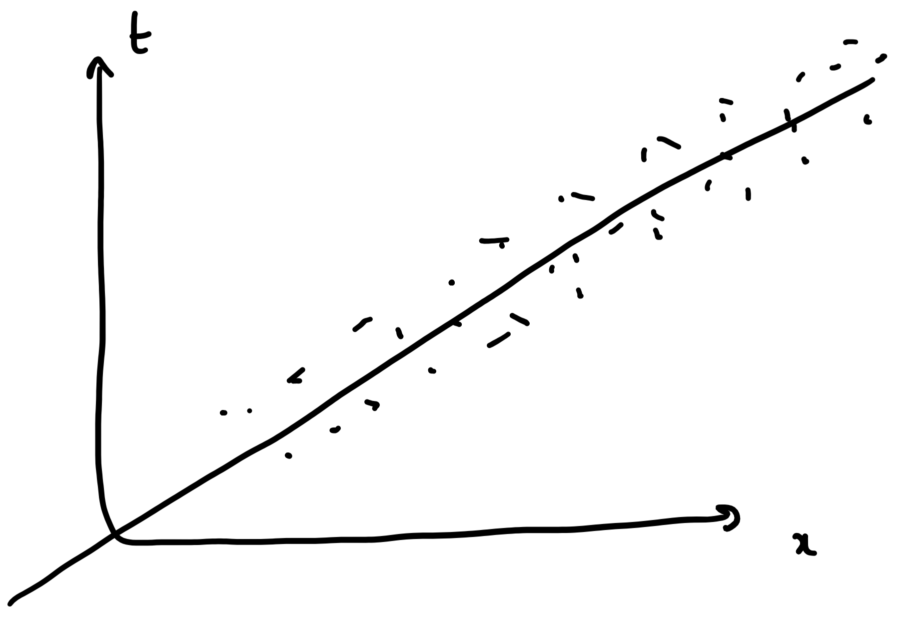
For entire dataset : N training examples

$$\boxed{\mathcal{E}(\underline{w,b}) = \frac{1}{2N}\sum_{i=1}^{N}\left(y^{(i)}-t^{(i)}\right)^2}$$

Why we are using this loss ?

$$t^{(i)} = \vec{w}^T\vec{x}^{(i)} + \underbrace{\Delta^{(i)}}_{\text{error term}}$$

is $\sim N(0,\sigma^2)$



Equivalent $t^{(i)} \sim N\left(\vec{w}^T\vec{x}^{(i)}, \sigma^2\right)$

|| Solution $\vec{w}^*$ to linear regression which maximizes likelihood estimate is the one that minimizes our least squared loss $\mathcal{E}(w,b)$

$$\vec{w}^* = \underset{\vec{w}}{\arg\max} \prod_{i=1}^{N} p\left(t^{(i)} \mid \vec{x}^{(i)}, \vec{w}\right)$$

if data points are independent.

1. We know that $x \to \ln(x)$ is monotonic
2. $\ln(AB) = \ln(A) + \ln(B)$

$$= \underset{\vec{w}}{\arg\max} \sum_{i=1}^{N} \ln\left(p\left(t^{(i)} \mid \vec{x}^{(i)}, \vec{w}\right)\right)$$

$$= \underset{\vec{w}}{\arg\max} \sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t^{(i)}-\vec{w}^T\vec{x}^{(i)}}{\sigma}\right)^2}\right)$$

PDF for $N(\mu,\sigma^2)$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$= \underset{\vec{w}}{\arg\max} \sum_{i=1}^{N} \underbrace{\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)}_{\substack{\text{is constant} \\ \text{(won't affect our maximization)}}} + \ln\left(e^{-\frac{1}{2}\left(\frac{t^{(i)}-\vec{w}^T\vec{x}^{(i)}}{\sigma}\right)^2}\right)$$

$$= \underset{\vec{w}}{\arg\max} \sum_{i=1}^{N} \ln\left(e^{-\frac{1}{2}\left(\frac{t^{(i)}-\vec{w}^T\vec{x}^{(i)}}{\sigma}\right)^2}\right)$$

ln is monotonic

$$= \underset{\vec{w}}{\arg\max} -\sum_{i=1}^{N} \frac{1}{2}\left(\frac{t^{(i)}-\vec{w}^T\vec{x}^{(i)}}{\sigma}\right)^2$$

maximizing $-x$ $\iff$ minimizing x

$$= \underset{\vec{w}}{\arg\min} \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(t^{(i)}-\vec{w}^T\vec{x}^{(i)}\right)^2$$

rescale

$\frac{1}{2\sigma^2} \longleftrightarrow \frac{1}{2N}$

$$= \underset{\vec{w}}{\arg\min} \frac{1}{2N}\sum_{i=1}^{N}\left(t^{(i)}-\vec{w}^T\vec{x}^{(i)}\right)^2$$

$$= \underset{\vec{w}}{\arg\min}\ \mathcal{E}(\vec{w})$$

This is the least squared loss from above.