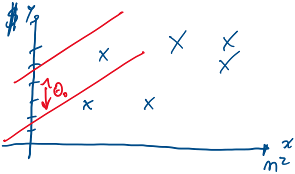


(x_1, y_1)

Collect n samples

$S = \{(x_1, y_1) \ldots (x_n, y_1)\}$

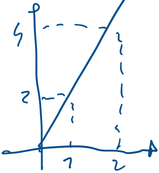
$h(x) \approx y$

$\downarrow$

$f(x) \approx y$

$S \cap T = \phi$

$p(x,y)$ is the same S&T



$\theta_0 = 0$

$\theta_1 = 2$

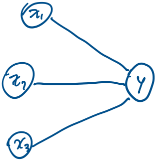
Identically distributed

Independently distributed :

$$f_\theta(x_i) = \theta_0 + \theta_1 x_i$$

$$x_i \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in R^m \quad y_i \in R$$

$$f_\theta(x^i) = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i \cdots \theta_m x_m^i$$



Friday, 18 October 2024 11:08

$p(x\, y)$

Training Set

$\downarrow$

Lear Algo

$\downarrow$

$$x \rightarrow \boxed{f} \rightarrow \hat{y}$$

$$f^* = \underset{f}{\operatorname{argmin}} \left\{ E_{p(x,y)} \left(\underbrace{\ell(y, f(x))}_{\text{Loss function}} \right) \right\}$$ Expected Risk

Monday, 14 October 2024
11:19

↑ high for bad approximations
↓ low for good ''

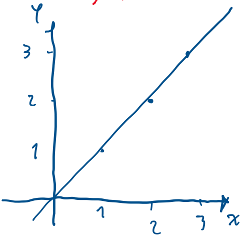
$\ell(\cdot)$ needs to be differentiable

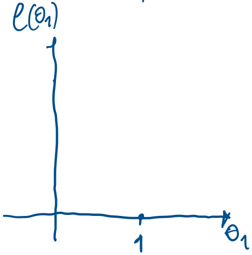
Empirical Risk Minimization

$$f^*(x) = \underset{f}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum^{n} \ell(y, f(x)) \right\}$$

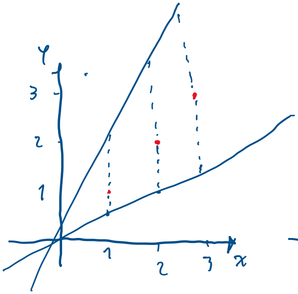
$$f(x) = \cancel{\theta_0} + \theta_1 x$$

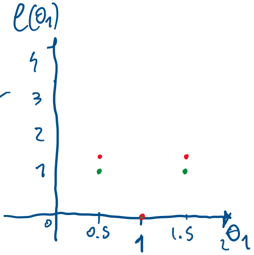
$J(\theta)$





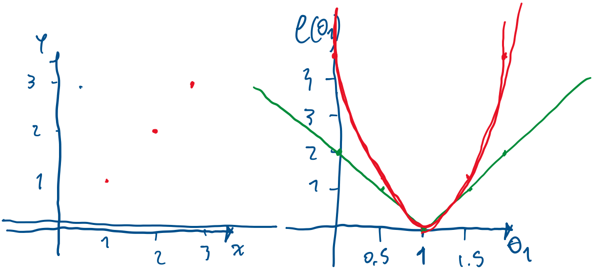
$y - f(x)$ squared loss : $(y - f(x))^2$
absolute : $|y - f(x)|$





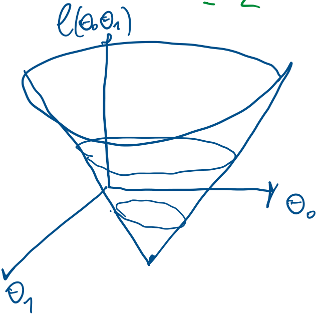
$$\frac{1}{3}\left((1-0.5)^2+(2-1)^2+(3-1.5)^2\right) \approx 1.2$$

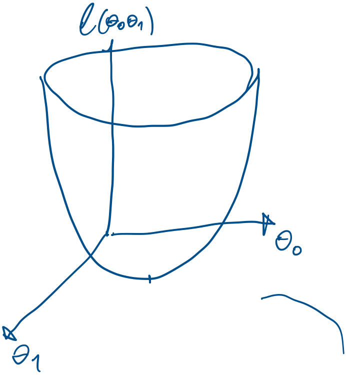
$$\frac{1}{3}\left(\overset{0.5}{|1-0.5|}+\overset{+\ 1}{(2-1)}+\overset{+\ 1.5}{(3-1.5)}\right] = 1$$

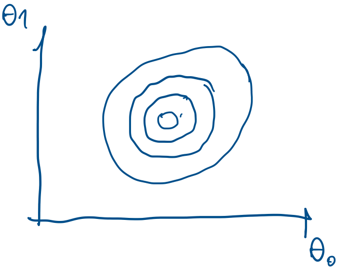


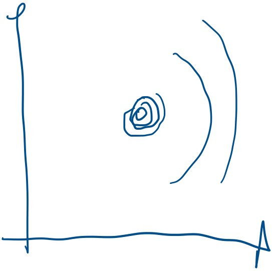
$$\frac{1}{3}\left(1^2+2^2+3^2\right) \approx 4.5$$

$$= 2$$





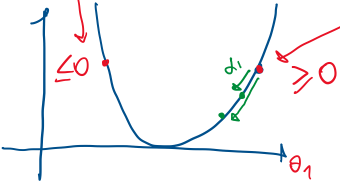


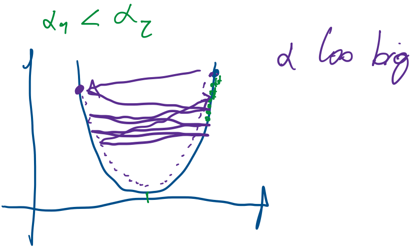


$\nabla \ell = \nabla J$

$$\nabla \ell = \frac{\partial \ell}{\partial \theta} = \left(\frac{\partial \ell}{\partial \theta_0}, \frac{\partial \ell}{\partial \theta_1}\right)$$

$$\theta_j \leftarrow \theta_j - \underbrace{\alpha}_{\text{learning Rate}} \frac{\partial}{\partial \theta_j}\left(\ell(\theta_0, \theta_1)\right) \quad \forall\, 0, 1$$





$(|f(x) - y|)^2$

as we approach the minimum
GD automatically takes smaller steps

$$\frac{\partial \ell(\theta_0, \theta_1)}{\partial \theta} = \frac{\partial \ell}{\partial f} \cdot \frac{\partial f}{\partial \theta}$$

$$\ell = \frac{1}{n} \sum^{n} \left(f_\theta(x_i) - y_i\right)^2$$

$$= \frac{1}{n} \sum^{n} \left(\theta_0 + \theta_1 x_i - y_i\right)^2$$

the loss depends on the prediction
and the prediction depends on the parameters

compute how loss change when prediction change

compute how the prediction change with respect to the params,

$$\frac{\partial \ell}{\partial f} = \frac{2}{n} \sum^{n} f(x_i) - y_i$$

$$\frac{\partial f}{\partial \theta_0} = \frac{\partial\, \theta_0 + \theta_1 x_i}{\partial \theta_0} = 1$$

$$\frac{\partial f}{\partial \theta_1} = x_i$$

$$\theta_0 \leftarrow \theta_0 - \alpha\left(\frac{2}{n}\sum^{n} f(x_i) - y_i\right)$$

$$\left[\theta_1 \leftarrow \theta_1 - \alpha\left(\frac{2}{n}\sum^{n}\left(f(x_i)-y_i\right)x_i\right.\right]$$

same if we have more paramenters to optimize

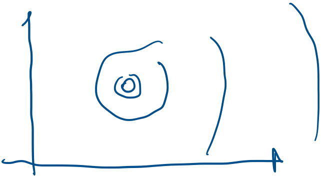
$\frac{1}{n}\sum^{n}$ = All train dataset : Batch GD

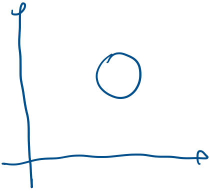
i = one sample : Stochastic GD

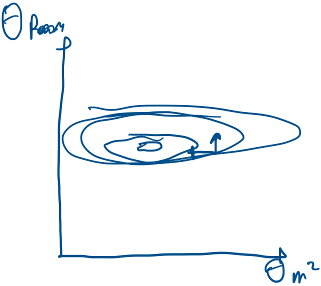
$\frac{1}{b}\sum^{b}$ = b samples : Mini-Batch GD

1 Iteration → compute G for one Mini Batch

1 Epoch $\frac{N}{b}$ iterations → one pass on train set







rooms 1~5

m² ≃ 100

$$p(x_1 y_1) = p(x_i)\, p(y_i|x_i)$$
$$= p(y_i)\, p(x_i|y_i)$$

$$p(S) = \prod_{i}^{n} p(x_i\, y_i)$$ Likelihood

$$\log p(S) = \log \sum_{i}^{n} p(x_i\, y_i)$$ Log Likelihood

model that approximates $p(x)$ $p(x|y)$

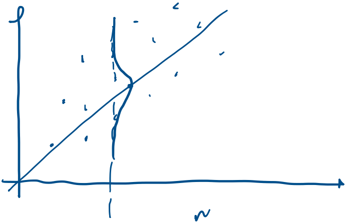
generative

model $p(y|x)$ we have a discriminative model

$$y^* = \underset{y}{\operatorname{argmax}}\; p(y|x)\, p(x)$$
$$= \underset{y}{\operatorname{argmax}}\; \log p(y|x) + \underbrace{\log p(x)}_{\text{does not depend on } y \text{ we can drop it}}$$

y^* = is the value that maximise the Log Likelihood of the conditional probability

lets use a Gaussian to model $p(y|x)$
lets use σ^2 fixed
lets use a function $f(x)$ to predict the mean of G.



$$p(y|x) = \mathcal{N}(y \mid f_\theta(x), \sigma^2)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y - f_\theta(x))^2}{\sigma}\right)$$

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \prod^{n} p(y_i \mid f(x_i), \sigma^2)$$

$$= \underset{\theta}{\operatorname{argmax}} \sum^{n} \underbrace{\log p(y_i \mid f(x_i), \sigma^2)}$$

forget about θ_0

↓

ias

$$= \underbrace{-\frac{1}{2}\log(\sigma) - \frac{1}{2}\log(2\pi)}_{\text{constant}} - \frac{1}{\sigma^2}(y_i - \theta_1 x_i)^2$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{\sigma^2}\sum^{N}(y_i - \theta_1 x_i)^2$$

$$\frac{1}{\sigma^2}\sum^{n}(y_i - f(x_i))^2 \quad = \frac{1}{n}\sum^{n}(y_i - f(x_i))^2$$

$$= -\log p(y_i \mid f(x))$$