Linear Regression (1/12/2023)

Predicting a real number

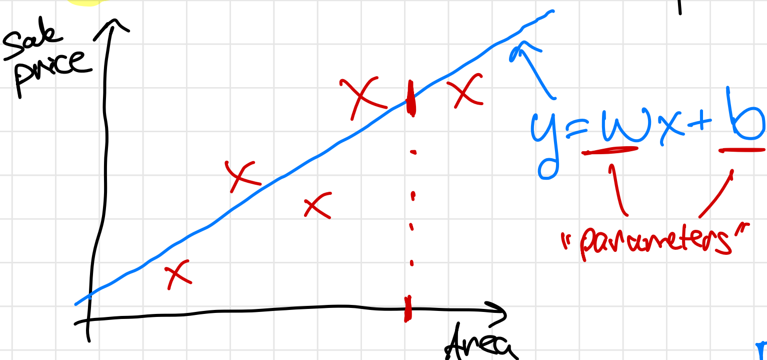
$n = |D|$

$D = \{(x^{(1)}, \boxed{y^{(1)}}), \ldots, (x^{(n)}, y^{(n)})\}$

$x^{(i)} \in \mathbb{R}^d$ $y^{(1)} \in \mathbb{R}$

"feature" (d total)

sale price	Area	#bedrooms	has garage?	bias
500K	1000	1	0	1
900K	1200	2	1	1



$$\hat{y} = \left(\sum_{i=1}^{d} w_i x_i\right) + b$$

$$= \underbrace{w}_{\text{parameters}}^{T} x + \underbrace{b}_{\text{parameters}}$$

$w \cdot x$

How to choose w & b?

Loss function

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}\Big(\underbrace{w^T x^{(i)} + b}_{\text{prediction}} - \underbrace{y^{(i)}}_{\text{true output "response"}}\Big)^2$$

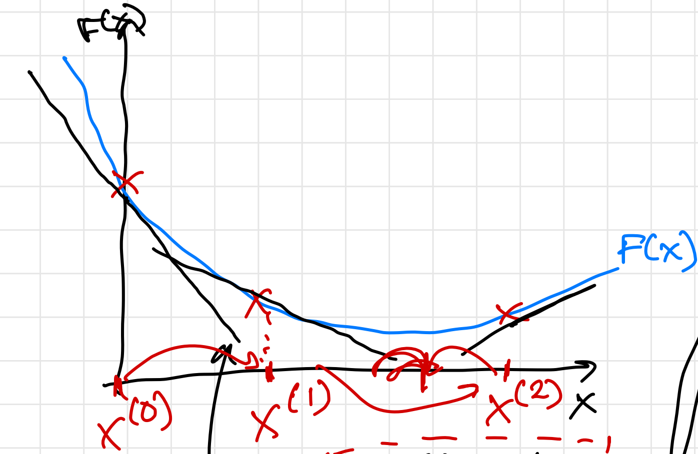
Find (w,b) that minimize $L(w,b)$

Optimization

Gradient Descent

Function F from $\mathbb{R}^d \to \mathbb{R}$, differentiable

Goal: minimize F



$$F(x) \approx F(x^{(0)}) + [x - x^{(0)}] \cdot F'(x^{(0)})$$

$$\nabla_x F(x) = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots \frac{\partial F}{\partial x_d} \right] \in \mathbb{R}^d$$

$$F(x) \approx F(x^{(0)}) + (x_1 - x_1^{(0)}) \nabla_x F(x^{(0)})_1 + (x_2 - x_2^{(0)}) \nabla_x F(x^{(0)})_2 \vdots (x_d - x_d^{(0)}) \nabla_x F(x^{(0)})_d$$

$$= \boxed{F(x^{(0)}) + (x - x^{(0)})^T \nabla_x F(x^{(0)})}$$

Fact: Gradient is direction of steepest ascent

negative gradient " " descent

Algorithm

$x^{(0)} \leftarrow \vec{0} \in \mathbb{R}^d$

For $t = 1, \dots, T$

$\quad x^{(t)} \leftarrow x^{(t-1)} - \boxed{\eta}\, \nabla_x F(x^{(t-1)})$

learning rate

return $x^{(T)}$

Let's start at $u \in \mathbb{R}^d$

and take a step $v \in \mathbb{R}^d$, $\|v\| \leq \varepsilon$

How to maximally increase $F(u+v)$?

$$\boxed{\|v\| = \sqrt{\sum_{j=1}^{d} v_j^2}}$$

$$F(u+v) \approx F(u) + (\cancel{u}+v-\cancel{u})^T \nabla F(u)$$

$$= F(w) + \underbrace{\|v\|}_{\leq \varepsilon} \cdot \underbrace{\|\nabla F(w)\|}_{\text{constant}} \cdot \underbrace{\cos(\alpha)}_{\substack{\text{Largest} = 1 \\ \text{Smallest} = -1}}$$

Announcements

$$\frac{d}{dw} 8w = 8$$

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(w^T x^{(i)} - y^{(i)}\right)}^{2}$$

$$\nabla_w L(w) = \underbrace{\frac{1}{n} \sum_{i=1}^{n}}\underbrace{2\left(w^T x^{(i)} - y^{(i)}\right)}_{\text{scalar}} \cdot \underbrace{x^{(i)}}$$

Algorithm for Linear Regression

$w^{(0)} \leftarrow \vec{0} \in \mathbb{R}^d$

for $t = 1, \ldots, T$

$\quad w^{(t)} \leftarrow w^{(t-1)} - \eta \cdot \frac{1}{n} \sum_{i=1}^{n} 2\left(w^{(t-1)T} x^{(i)} - y^{(i)}\right) \cdot x^{(i)}$

return $w^{(T)}$