V2 Sammanfatta linjär regression (allmän linjär regression)

Gradient descent, maximum likelihood

Regaljärisering

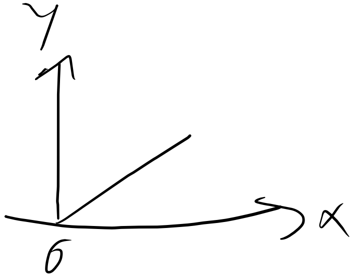
Korsvalidering

Allmän linjär regression

$y = f(x)$ $\qquad \varphi(x, \cancel{y}) = (y, f(x))$

Vi söker linjära approximation givet ett stickprov (träningsdata)

(local regression, kap 7)



$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Y	X
Y_1	x_1
Y_2	x_1
$\vdots$	$\vdots$

Y	X_0	X_1
Y_1	1	x_a
Y_2	1	x_b
		$\vdots$

β_0, β_1 skall bestämmas

$$\min \mathbb{C}(\mathbb{B}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

Hur fungerar det?

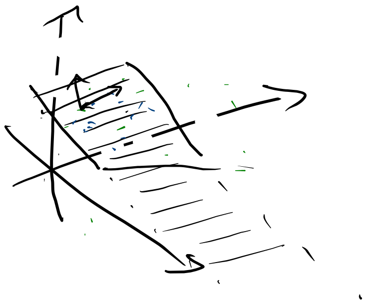
$$(\text{optimering}) \approx \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2$$

· Derivata → sätt till 0 → analys

ϵ^2 står kvar i högerledet, dvs vi optimerar mot låg <u>varians</u>.

$$E[Y|X] = \mathbb{X}\mathbb{B}$$

3p:



"Lutning på plan" = gradient

$\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}\right) = \Delta f(x_1, x_2)$

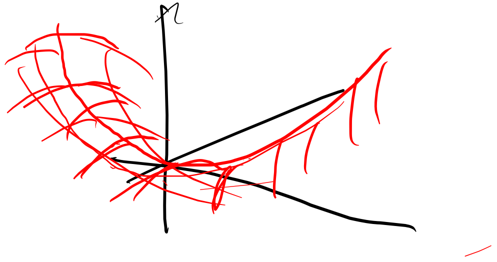
gradienten pekar alltid i den riktning där förändringshastigheten är störst.

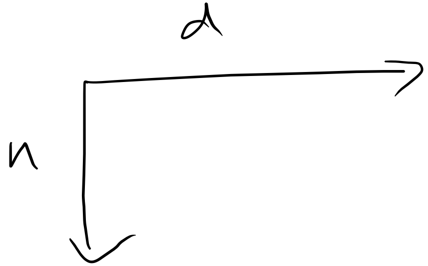
$Y = X\beta + E$

Feature-expansion för att lämna strikta linjära förhållanden

$Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \beta_3 x_1 x_2$

$Y = \beta_1 x_1 x_2$





$O(d^2 n)$

$E[Y|X] = f(X)$

Polynom-expansion:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 y^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

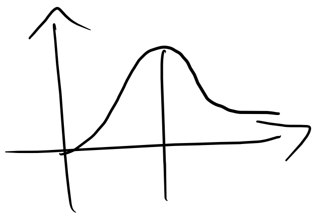
Generell Additiv Modell

(kap 7 i boken)

$$Y = f_0(X_0) + f_1(X_1) + \dots + f_d(X_d) \quad (+\epsilon)$$

$$\Rightarrow \sum_{i=1}^{n} x_0 f_0(X_0) \Rightarrow \underline{x_0 f(X)}$$

$$\sim \underline{\Pr(Y \mid X)}$$



$$E[X] = \mu$$

länk funktion för linjär reg. : μ

$MSE = Bias^2 + varians + brus$ (irreducibel)

reguljärisering minskar varians, men tenderar att öka bias.

underfit: (för enkel modell)

låg varians, högt bias

(hög precision, låg noggranhet)

overfit: (för komplicerad)

hög varians, lågt bias

(låg precision, hög noggranhet)

Reguljärisering

$$\min \mathbb{C}(\beta) = RSS + \lambda \sum_{i=1}^{d} \beta_d^2$$ l_2 - norm

- Ridge Regression

$$\min \mathbb{C}(\beta) = RSS + \lambda \sum_{i=1}^{d} |\beta_d|$$ l_1 - norm

- LASSO regression

$$\min \mathbb{C}(\beta) = RSS + \lambda \left(\frac{1-\alpha}{2} \sum_{i=1}^{d} \beta_i^2 + \alpha \sum_{i=1}^{d} |\beta_i| \right)$$

- Elasticnet regression