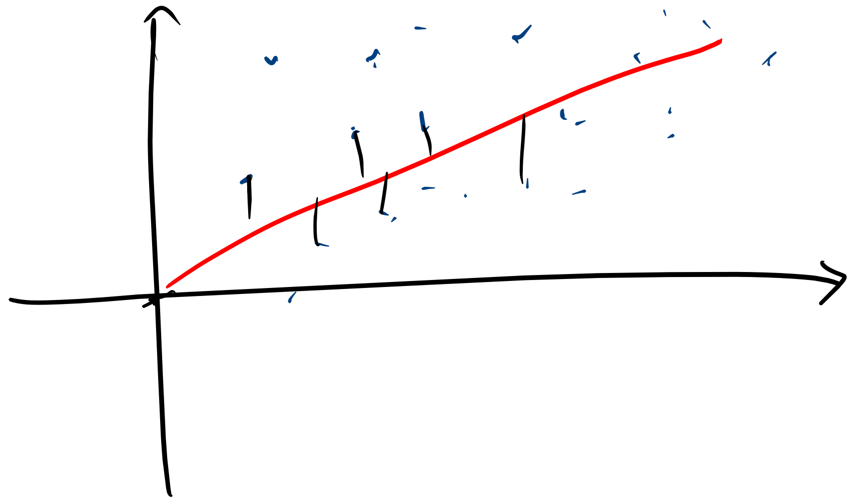
Skattning (Estimation)

$\mathbb{E}[X] = \mu \Rightarrow \bar{X}$ ↙ Skattare (estimators) $\hat{\mu}$

$\operatorname{Var} X = \sigma^2 \Rightarrow s^2$ ↙ $\hat{\sigma}$ (rättvisa skattare)



Felen runt regressionslinjen kallas (kvadrerade) SSE (sum of square errors)

En regression vill minimera SSE!

$$SSE = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

$$\frac{\partial SSE}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial SSE}{b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i = 0$$

$$nb_0 + b_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_0\sum_{i=1}^{n} x_i + b_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\underline{b = (A^T A)^{-1} A^T y}$$

$$
\begin{array}{cc}
\beta_0 & \beta_1 \\
\end{array}
$$

$$
X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & \vdots \end{pmatrix}
$$

$$
Y = X\beta + E \iff Y = \beta_0 + \beta_1 X + E
$$

$$
\frac{d}{dx} kx + m = k \iff \frac{d}{dx} kx^1 + m = 1 \cdot k x^0 + 0 = k
$$

Power law for derivatives

$$
\frac{d}{dx} x^a = a x^{a-1}
$$

När vi gjort regression så har vi en skattning av medlet i Y givet X

$$\mu_{Y|x} = \beta_0 + \beta_1 X$$

Test statistikor

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

För hypoteser:

$H_0: \beta_1 = 0$

$H_1: \beta_1 > 0$

Högersidigt test (right-tailed)

$H_0: \beta_1 = 0$

$H_1: \beta_1 < 0$

Vänstersidigt test (left-tailed)

"Inference on slope" (inferens på lutningen)

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$$T_{n-2} = \frac{B_1}{s/\sqrt{S_{xx}}}$$

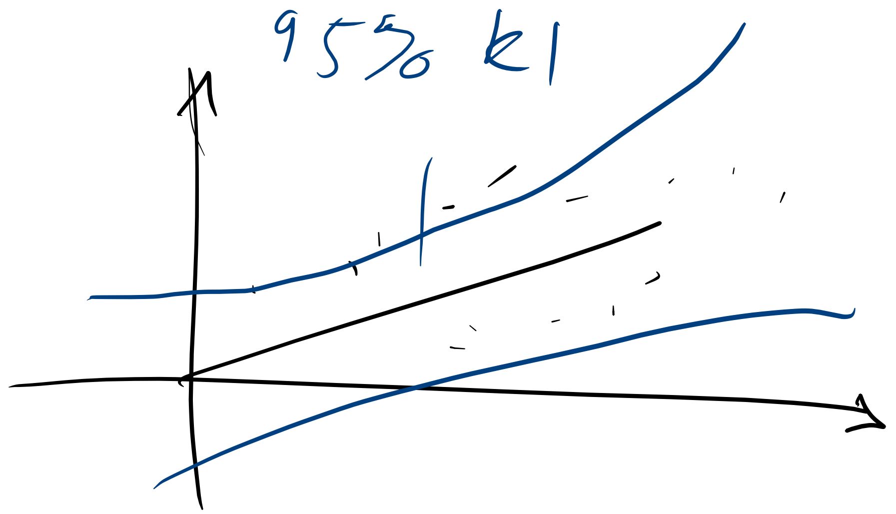
S = stickprovs avvikelsen

S_{xx} = Variabilitet i X

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Konfidens intervall:

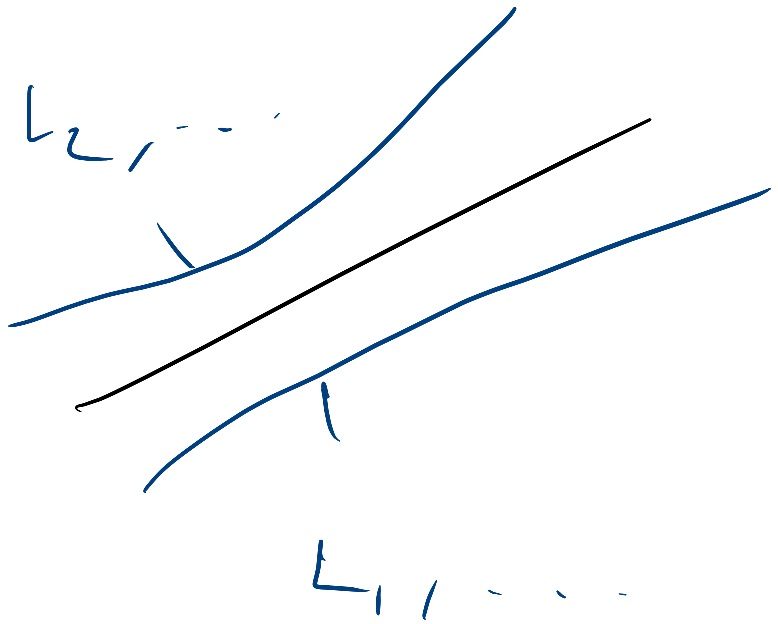
$$B_1 \pm t_{\alpha/2} \, S \sqrt{S_{xx}}$$



$$\hat{\mu}_{Y|x} \pm t_{\alpha/2}\, s \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

Räkna ut ovanstående för hela datamängden:

$[L_1 \;,\; L_2]$



Vårde regression: förutsäg ett numeriskt värde.
Klassificering: förutsäg en egenskap

Maximum Likelihood Estimation

$$\hat{\theta} = \operatorname{argmax} \mathcal{L}(\theta; y)$$

För n gemensamma fördelningar (deras parametrar θ)

$$f_n(y \mid \theta) = \prod_{k=1}^{n} f_k(y_k; \theta) \Rightarrow \mathcal{L}(\theta; y)$$

detta är ett maximeringsproblem, en lossfunktion är dock ett minimeringsproblem:

$$l_k = -y_k \ln p_k - (1-y_k) \ln (1-p_k)$$

$$L = \prod_{k: y_k = 1} p_k \prod_{k: y_k = 0} (1 - p_k)$$

För komplicerade gemensamma fördelningar kan vi inte alltid hitta vettiga formler för likelihood-funktioner.

Då vänder vi oss i ML till

Stochastic Gradient Descent