IM 12.5.09

Notiztitel 12.05.2009

Lineare Regression

$$F = \sum_{i=1}^{n} \left(x^{(i)} - \sum_{j=0}^{k} w_j \cdot a_j^{(i)} \right)^2 \overset{!}{=} \min$$

$$\frac{\partial F}{\partial w_\ell} = \sum_{i=1}^{m} 2 \cdot \left(x^{(i)} - \sum_{j=0}^{k} w_j \cdot a_j^{(i)} \right) \cdot \left(-a_\ell^{(i)} \right) \overset{!}{=} 0 \text{ für } \ell = 0 \ldots k$$

$$\Leftrightarrow \sum_{i=1}^{m} \sum_{j=0}^{k} \underline{w_j} \cdot a_j^{(i)} a_\ell^{(i)} = \sum_{i=1}^{m} a_\ell^{(i)} x^{(i)} \text{ für } \ell = 0 \ldots k$$

Lineares GLS, nach den w_j auflösen

Lineare Regression zur Klassifikation

Regressionsgerade

Fehler der Regression, die keine echten Fehler sind

Paarweise Regression

Klassen A, B, C

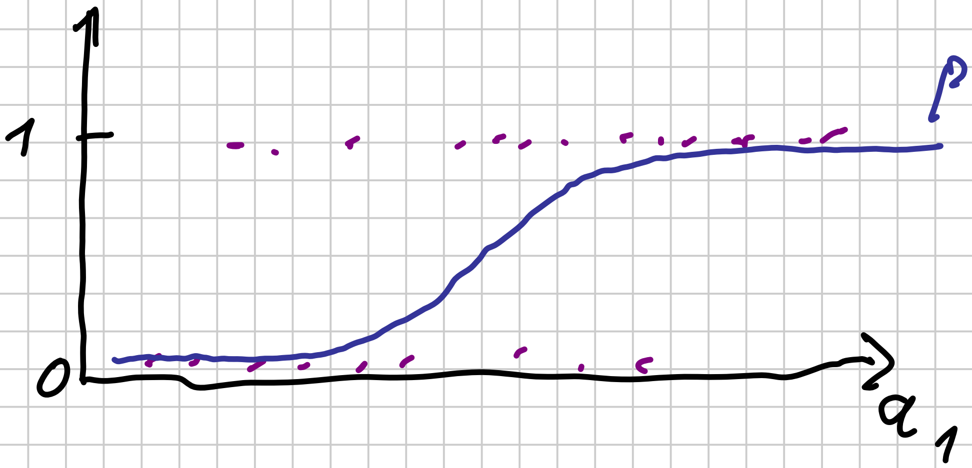
A : B : −0,8
A : C : +0,2
B : C : +0,7

Abstimmen für eine Instanz

A : −0,6
B : +1,5
C : −0,9

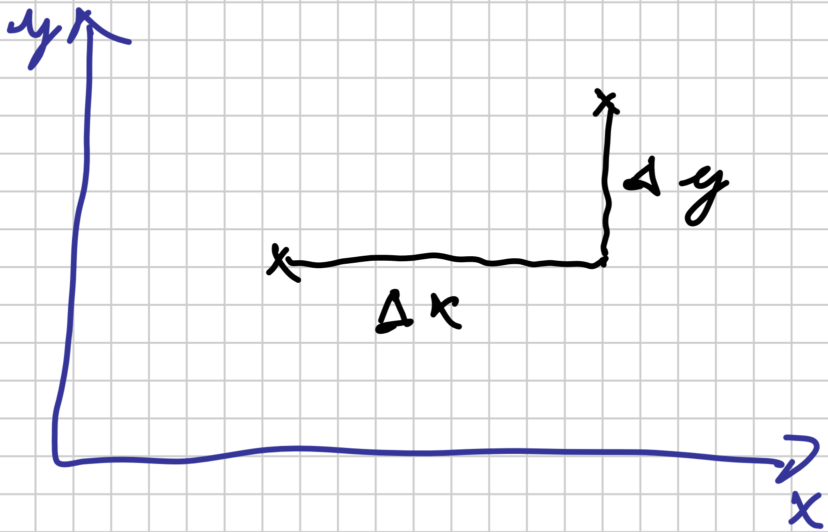
→ wähle Klasse B

Logistische Regression



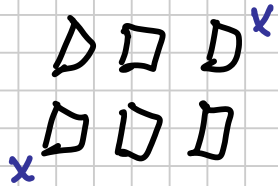
$$p \approx \frac{e^{w_0 + w_1 a_1}}{1 + e^{w_0 + w_1 a_1}}$$

Distanz-Metriken



Euklidische Distanz $\sqrt{(\Delta x)^2 + (\Delta y)^2}$

Manhattan-Distanz $|\Delta x| + |\Delta y|$



Bayessches Netzwerk

$P(play, outlook, windy, temperature, humidity) =$

naiver Bayes:

$= P(outlook \mid play) \cdot P(windy \mid play) \cdot P(temperature \mid play) \cdot P(humidity \mid play) \cdot P(play)$

Netzwerk

$= P(outlook \mid play) \cdot P(windy \mid outlook, play) \cdot P(temperature \mid outlook, play) \cdot P(humidity \mid temperature, play) \cdot P(play)$

$P(windy = true, outlook = sunny, temperature = cool, humidity = high \mid play = yes) =$

$P(outlook = sunny \mid play = yes) \cdot P(windy = true \mid outlook = sunny, pl = yes) \cdot$

$P(temperature = cool \mid outlook = sunny, play = yes) \cdot$

$P(humidity = high \mid temperature = cool, play = yes) \cdot P(play = yes)$