Formelsammlung

Evaluation

Precision

$$P = \frac{tp}{tp + fp}$$

Recall

$$P = \frac{tp}{tp + fn}$$

F-Score

$$F_\beta = (1 + \beta^2)\frac{PR}{\beta^2 P + R}$$

Vereinfachung für $\beta = 1$

$$F_1 = 2\frac{PR}{P + R}$$

Wahrscheinlichkeitsrechnung und Bayes' Law

Definition bedingte Wahrscheinlichkeit

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

Kettenregel

$$\begin{aligned} p(A,B,C) &= p(A|B,C)p(B,C) \\ &= p(A|B,C)p(B|C)p(C) \end{aligned}$$

Bayes'sches Gesetz

$$\begin{aligned} p(B|A) &= \frac{p(A,B)}{p(A)} \\ &= \frac{p(A|B)p(B)}{p(A)} \end{aligned}$$

Inter-Annotator Agreement (Fleiss' Kappa)

Parameter

N	Anzahl der annotierten Instanzen
k	Anzahl der annotierbaren Kategorien
p_j	Wahrscheinlichkeit für die Wahl von Kategorie j
O_i	Beobachtete Übereinstimmung für Instanz i

Beobachtete Übereinstimmung (observed agreement)

$$O = \frac{1}{N}\sum_{i=1}^{N} O_i$$

Erwartete Übereinstimmung (expected agreement)

$$E = \sum_{j=1}^{k} p_j^2$$

Fleiss' Kappa

$$\kappa = \frac{O - E}{1 - E}$$

Lineare Algebrea / Vektoren

Skalarprodukt

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{N} a_i b_i$$

Länge eines Vektors

$$|\vec{a}| = \sqrt{\sum_{i=1}^{N} a_i^2}$$

Kosinus-Ähnlichkeit

$$\begin{aligned} s(\vec{a}, \vec{b}) &= \frac{\sum_{i=1}^{N} a_i b_i}{\sum_{i=1}^{N} a_i^2 \sum_{i=1}^{N} b_i^2} \\ &= \frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|} \end{aligned}$$

Regression und neuronale Netze

Sigmoid-Funktion

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Loss-Funktion Mean Squared Error
(w: Gewichte, m: Anzahl der Instanzen im Datensatz, x_i und y_i: Ein- bzw. Ausgabewerte von Instanz i, h: Hypothesenfunktion)

$$J(\vec{w}) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_{\vec{w}}(x_i) - y_i\right)^2$$

Loss-Funktion Binary Cross Entropy
(w: Gewichte, m: Anzahl der Instanzen im Datensatz, x_i und y_i: Ein- bzw. Ausgabewerte von Instanz i, h: Hypothesenfunktion)

$$J(\vec{w}) = -\frac{1}{m} \sum_{i=0}^{m} \Big(\big(y_i \log h_{\vec{w}}(x_i)\big) + \big((1 - y_i) \log(1 - h_{\vec{w}}(x_i))\big) \Big)$$