Formulario

Funções de Distância

Distância Euclediana $$dist(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2} \quad (1)$$

Distância Minkowski $$dist(x,y) = \sqrt[m]{|x_1 - y_1|^m + |x_2 - y_2|^m + ... + |x_n - y_n|^m} \quad (2)$$

Distância Manhattan $$dist(x,y) = |x_1 - y_1| + |x_2 - y_2| + ... + |x_n - y_n| \quad (3)$$

Distância de Hamming $$dist(x,y) = \sum_{i=1}^{n} x_i \neq y_i \quad (4)$$

Avaliação de Modelos

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Tx.Erro = 1 - accuracy \quad (6)$$

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

$$TNR = \frac{TN}{TN + FP} \quad (8)$$

$$FPR(fallout) = \frac{FP}{FP + TN} \quad (9)$$

$$precision = \frac{TP}{TP + FP} \quad (10)$$

$$recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (12)$$

Probabilidade Condicional Teorema de Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)} \tag{13}$$

$$V_{MPP} = \arg max_{v_j \in V} P(v_j) \times \prod P(A_i|v_j) \tag{14}$$

Atributos contínuos $$P(A_i|c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{\frac{-(A_i - \mu ij)^2}{2\sigma_{ij}^2}} \tag{15}$$

Regras de Associação

$$Sup(A \Rightarrow B) = \frac{P(A \cap B)}{n} \tag{16}$$

$$Conf(A \Rightarrow B) = \frac{P(A \cap B)}{P(A)} \tag{17}$$

$$Cob(A \Rightarrow B) = \frac{Sup(A \Rightarrow B)}{Conf(A \Rightarrow B)} \tag{18}$$

$$Interesse(A \Rightarrow B) = \frac{Sup(A \Rightarrow B)}{Sup(A) \times Sup(B)} \tag{19}$$

$$Leverage(A \Rightarrow B) = Sup(A \Rightarrow B) - Sup(A) \times Sup(B) \tag{20}$$

$$Convinction(A \Rightarrow B) = \frac{1 - Sup(B)}{1 - Conf(A \Rightarrow B)} \tag{21}$$

Text Mining

$$TF - IDF = \frac{N_t}{N_d} \times log_{10}(\frac{N_{dc}}{N_{dt}}) \tag{22}$$