2022.03.02

Naive Bayes

$P(X \cap Y) = P(X|Y)P(Y) = P(Y|X)P(X)$

一般已知其余3项求 $P(X|Y)$ 或 $P(Y|X)$

Given a record with attribute ($A_1, A_2 \cdots$)

: Goal: predict 其class，找C such that $P(C|A_1, A_2 \cdots A_n)$ 最大

$$\underset{\text{maximize}}{P(C|A_1, A_2 \cdots A_n)} = \frac{P(A_1, \cdots A_n|C)\, P(C)}{P(A_1 \cdots A_n)}$$

为已知，整个式子取决于 $P(A_1 \cdots A_n|C)$

使 $P(A_1 \cdots A_n|C)$ 最大：

Assume attributes 间 independent，则

$P(A_1 \cdots A_n|C) = P(A_1|C) \cdot P(A_2|C) \cdots P(A_n|C)$

$P(A_1|C)$. 看C那个class中有多少 attribute A_1 为此值的.

Discrete Attribute 好算，但是 continuous 需一些方法.

① Discretize

分range 等.

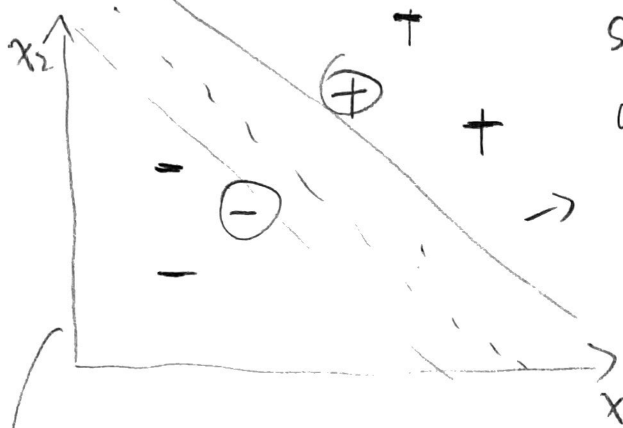
② 将 A_1 的 distribution 看作 normal.

算所有class的 $P(A_1, \cdots A_n|C)$ 取最大

· What if $P(A_i|C) = 0$ for some i?

Laplace: $P(A_i|C) = \frac{N_{ic}+1}{N_c+c}$ c为class数.

Support Vector Machine



SVM找widest street，street最中间为分离data的boundary line.

找出，line应如何分？

例：line: $w_1x_1 + w_2x_2 + b = 0$

(w_1, w_2) 为1个垂直于该线的vector

$\vec{w}\cdot\vec{u}+b \geq 0$ 则 +，<0 则 -

只有标出的2个data point有助于找此street 称support vector.

· How to find the widest street.

samples需全在street外

$\vec{w}\cdot\vec{x}_+ + b \geq 1$ $\qquad$ $\vec{w}\cdot\vec{x}_- + b \leq 1$

"+"需 $\qquad$ "-"需

improved method

$y_i(\vec{w}\cdot\vec{x}_i+b)-1=0$ $\qquad$ $y_i = \begin{cases}1 & \text{if } + \\ -1 & \text{if } -\end{cases}$

求width（用support vector）

求$\vec{x}_+$，$\vec{x}_-$在$\vec{w}$上的projection，再算proj magnitude差.

width $= \frac{2}{|\vec{w}|}$

令此maximum，即minimize $|\vec{w}|$

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_i \alpha_i [y_i(\vec{x}_i\cdot\vec{w}+b)-1]$$

不在boundary上的 x_i 的 $\alpha_i = 0$

无法 linearly separable data?

① 允许 misclassification，仍用 line 分.

② Transformation.

Polynomial kernal

$$K(\vec{x_i},\vec{x_j}) = (\vec{x_i}\cdot\vec{x_j}+1)^n$$

Radical

$$K(\vec{x_i},\vec{x_j}) = e^{\frac{\|\vec{x_i}-\vec{x_j}\|}{\sigma}}$$

将 L 的 derivative 中的 $\vec{x_i}\cdot\vec{x_j}$ 替代为 Kernel function