Bayesian ML.

1. Naive Bayes.

$$p(C_k|\vec{x}) = \frac{p(C_k)\cdot p(\vec{x}|C_k)}{p(\vec{x})} \propto p(C_k, \vec{x}) =$$

$$= p(x_1|x_2, \dots x_n, C_k)\cdot p(x_2|x_3, \dots x_n, C_k)\cdot \dots \cdot p(x_n|C_k)\, p(C_k) =$$

$$= \Big(\prod_{i=1}^{n} p(x_i|C_k)\Big)\, p(C_k)$$

Naiveness $\Downarrow$

$$\hat{y}(\vec{x}) = \underset{k\in\{1,\dots k\}}{\operatorname{argmax}} \Big(p(C_k) \prod_{i=1}^{n} p(x_i|C_k)\Big)$$

1) $p(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\Big\{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}\Big\}$ – Gaussian NB (numerical features)

2) $p(\vec{x}|C_k) = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n}(x_i!)}\prod_{i=1}^{n} p_{ki}^{x_i}$, p_{ki} – probab. that event i occurs in class k.

x_i – number of occur. for event i

Multinomial NB (categorical features)

3) $p(\vec{x}|C_k) = \prod_{i=1}^{n} p_{ki}^{x_i}\cdot(1-p_{ki})^{(1-x_i)}$, p_{ki} – –//–,

x_i – binary label if event i occured.

Example ~~NB + Logis~~ TF-IDF + NB + Logistic Regression (Wang, 2012) for sentiment analysis.

1) X – texts, y – ~~la~~ binary labels

$\vec{x}$ = tf-idf(X, ngram_range=(1,2)).

~~2)~~ $\cancel{p(C_0) =}$

$$\cancel{\frac{p(C_1|\vec{x})}{p(C_0|\vec{x})} = \frac{p(C_1)\cdot\prod_{i=1}^{n} p_{1i}^{x_i}}{p(C_0)\prod_{i=1}^{n} p_{0i}^{x_i}}}$$

2) $p(r_1|\vec{x}) > p(r_0|\vec{x})$

$$\Updownarrow$$

$$\cancel{\log\left\{\frac{p(r_1|\vec{x})}{p(r_0|\vec{x})}\right\} > 0}$$

$$\cancel{\Updownarrow}$$

$$\cancel{\vec{w}^T\vec{x} + b > 0, \text{ where } w_i = \log\left(\frac{p_{1i}}{p_{0i}}\right),\ b = \log\left(\frac{p(r_1)}{p(r_0)}\right).}$$

~~linear~~

$$\cancel{\log\left(\frac{p(r_1|\vec{x})}{1-p(r_1|\vec{x})}\right) > 0}$$

$$\Updownarrow$$

$$p(r_1|\vec{x}) > 0.5$$

with MNB: $\cancel{p(r_1|\vec{x}) = p(r_1)\cdot\pi},\ \cancel{\pi = \frac{(\sum_{i=1}^{n} x_i)!}{(\prod_{i=1}^{n} x_i!)\cdot p(\vec{x})}}$

$$p(r_1|\vec{x}) = p(r_1)\cdot\pi\cdot\prod_{i=1}^{n} p_{1i}^{x_i},\quad \pi = \frac{(\sum_{i=1}^{n} x_i)!}{(\prod_{i=1}^{n} x_i!)\cdot p(\vec{x})}$$ - hard to compute.

<u>idea</u>: reexpress $p(r_1|\vec{x})$ as the output of logistic regression:

$$p(r_1|\vec{x}) = \frac{1}{1+e^{-z}} \Leftrightarrow z = \ln\left(\frac{p(r_1|\vec{x})}{1-p(r_1|\vec{x})}\right) = \ln\left(\frac{p(r_1|\vec{x})}{p(r_0|\vec{x})}\right) =$$

$$= \ln\left(\frac{p(r_1)}{p(r_0)}\right) + \sum_{i=1}^{n}\ln\left(\frac{p_{1i}}{p_{0i}}\right)\cdot x_i$$

p_{1i} is estimated as: $p_{1i} = \frac{\sum_{j: y^{(j)}=1} x_i^{(j)} + 1}{\ldots}$ same with p_{0i},

afterwards we apply logistic regression to $\tilde{x} = \left\{\ln\left(\frac{p_{1i}}{p_{0i}}\right)\cdot x_i\right\}_{i=0}^{n}$.