From Bayes' theorem to Softmax

x.li 20200225 livvddcc@gmail.com

We use Bayes' theorem to calculate the conditional probability P(C1|x):

$$\begin{aligned} P(C_1|x) &= \frac{P(C_1,x)}{P(x)} \\ &= \frac{P(x|C_1)P(C_1)}{p(x|C_1)P(C_1)+P(x|C_2)P(C_2)} \\ &= \frac{1}{1+\frac{P(x|C_2)P(C_2)}{p(x|C_1)P(C_1)}} \\ &= \frac{1}{1+e^{-a}} \qquad [sigmoid] \end{aligned}$$

Where:

$$a = In\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$$

We use Gaussian distribution as maximum likelihood estimation:

$$P(x|C_1) \sim N(x|\mu_1,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\}$$

$$P(x|C_2) \sim N(x|\mu_2,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}exp\{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\}$$

$$InP(x|C_1) = -\frac{D}{2}In(2\pi) - \frac{1}{2}In|\Sigma| - \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

$$InP(x|C_2) = -\frac{D}{2}In(2\pi) - \frac{1}{2}In|\Sigma| - \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)$$

So we get the sigmoid function:

$$\begin{aligned} a(x) &= InP(x|C_1) - InP(x|C_2) + In\frac{P(C_1)}{P(C_2)} \\ &= (\mu_1-\mu_2)^T\Sigma^{-1}x - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + In\frac{P(C_1)}{P(C_2)} \\ &= w^Tx + w_0 \end{aligned}$$

Where:

$$w = \Sigma^{-1}(\mu_1-\mu_2)$$

$$w_0 = \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + In\frac{P(C_1)}{P(C_2)}$$

$$P(C_2|x) = 1 - P(C_1|x)$$

The sigmoid function is used for the two-class logistic regression, whereas the softmax function is used for the multiclass logistic regression.

$$\begin{aligned} P(C_k|x) &= \frac{P(x|C_k)P(C_k)}{\sum_j P(x|C_j)P(C_j)} \\ &= \frac{exp(a_k)}{\sum_j exp(a_j)} \end{aligned}$$