Machine Learning homework 4 solution

Linear Classification

Wiktor Jurasz - M.Nr. 03709419

November 18, 2018

1 Problem 1

a Exponential distribution

b For binary generative models posterior is given by

$$p(y = k_1|x) = \frac{p(x|y = k_1)p(y = k_1)}{p(x|y = k_1)p(y = k_1) + p(x|y = k_0)p(y = k_0)} \tag{1}$$

If we plug in given prior and conditional we get for $y = 1$

$$p(y = 1|x) = \frac{\frac{1}{2}\lambda_1 e^{-\lambda_1 x}}{\frac{1}{2}\lambda_1 e^{-\lambda_1 x} + \frac{1}{2}\lambda_0 e^{-\lambda_0 x}} \tag{2}$$

and similar for $y = 0$.

With given classification decision, to check for which values x has class $y = 1$ we have to check for which x

$$p(y = 1|x) > p(y = 0|x) \tag{3}$$

Because when this inequity is true, then our argmax_k will return $y = 1$.
As the denominator is the same and $\frac{1}{2}$ cancels out, we have:

$$\lambda_1 e^{-\lambda_1 x} > \lambda_0 e^{-\lambda_0 x} \Leftrightarrow \ln \lambda_1 e^{-\lambda_1 x} > ln\lambda_0 e^{-\lambda_0 x} \Leftrightarrow \ln \lambda_1 - \lambda_1 x > \ln \lambda_0 - \lambda_0 x \Leftrightarrow x < \frac{\ln \lambda_1 - \ln \lambda_0}{\lambda_1 - \lambda_0} \tag{4}$$

2 Problem 2

Maximum likelihood solution for logistic regression is given when $\sigma = 0.5$. This corresponds to $w^T\phi = 0$ when magnitude of w goes to infinity. In consequence sigmoid function becomes extremely steep (Heaviside function) and every point x has probability of 1 to belong to its class. This leads to heavy overfitting. To solve this issue we can (the same as in logistic regression) use regularization term $\lambda||w||_q^2$ in error function.

3 Problem 3

$$softmax = S(x)_i = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \tag{5}$$

For $K = 2$:

$$S(x)_0 = \frac{e^{x_0}}{e^{x_0} + e^{x_1}} = \frac{\frac{e^{x_0}}{e^{x_0}}}{\frac{e^{x_0}}{e^{x_0}} + \frac{e^{x_1}}{e^{x_0}}} = \frac{1}{1 + \frac{e^{x_1}}{e^{x_0}}} = \frac{1}{1 + e^{\ln \frac{e^{x_1}}{e^{x_0}}}} = \frac{1}{1 + e^{-\ln \frac{e^{x_0}}{e^{x_1}}}} = \sigma(\ln \frac{e^{x_0}}{e^{x_1}}) \tag{6}$$

4 Problem 4

One example of such a function could be:

$$\phi(x_1, x_2) = (y, x) = (x_1 x_2, x_1) \tag{7}$$

In this case all $y's$ are positive for circles and negative for crosses. x_1 stays the same and it is possible to draw a line around $y = 0$ to separate both sets.