Week 3 Homework Submission
Linear Classification

Iuliia Skobleva
Matriculation Number 03723809

November 17, 2019

Linear Classification

Problem 1

We have a uniform prior $p(y = 0) = p(y = 1) = 0.5$ on binary labels y and exponential distributions as class conditionals $p(x|y = 0) = \text{Expo}(x|\lambda_0)$ and $p(x|y = 1) = \text{Expo}(x|\lambda_1)$, where λ_0 and λ_1 are known and fixed. The posterior distribution for a 2-class case is a Bernoulli distribution.

The next question is what values of x are classified as class 1? In class we found:

$$p(y = 1|x) = \sigma(a) \tag{1}$$

where $\sigma(a)$ is a sigmoid function and $a = \ln \frac{p(x|y=1)p(y=1)}{p(x|y=0)p(y=0)}$. In our case:

$$\begin{aligned} a &= \ln \frac{\lambda_1 \exp -\lambda_1 x}{\lambda_0 \exp -\lambda_0 x} \\ &= \ln \frac{\lambda_1}{\lambda_0} + \ln \frac{\exp -\lambda_1 x}{\exp -\lambda_0 x} \\ &= \ln \frac{\lambda_1}{\lambda_0} + x(\lambda_0 - \lambda_1) \end{aligned} \tag{2}$$

We also know that the decision boundary of a sigmoid function lies by $a = 0$ and, here, if $a > 0$ then values will be classified as class 1. This means, if $x > -\frac{1}{\lambda_0 - \lambda_1} \ln \frac{\lambda_1}{\lambda_0}$, all instances are classified as 1.

Problem 2

For a linearly separable dataset the decision boundary occurs at $\sigma(w^T x) = 0.5$ or $w^T x = 0$. If the weights w become very large (which is what happens when we only consider the training data) the sigmoid function essentially becomes the Heaviside-stepfunction and every training point of class k is assigned to this class k with probability 1, which can poorly generalize to the new data. The problem of large weights can be avoided, the same way as in linear regression, by penalizing large weights using a regularization term.

Problem 3

The definition of a softmax is:

$$\sigma(\mathbf{x})_i = \frac{\exp(x_i)}{\sum_{k=1}^{K} \exp(x_k)} \tag{3}$$

where K is the number of classes and $\mathbf{x}$ is the posterior probability vector:

$$\mathbf{x} = \begin{pmatrix} p(y=0|x) \\ p(y=1|x) \end{pmatrix} \tag{4}$$

Let's look at the first component of this vector:

$$\begin{aligned} \sigma(p(y=0|x)) &= \frac{\exp p(y=0|x)}{\exp p(y=0|x) + \exp p(y=1|x)} \\ &= \frac{1}{1 + \exp -[p(y=0|x) - p(y=1|x)]} \\ &= \sigma_{sig}(a) \end{aligned} \tag{5}$$

where $a = p(y=0|x) - p(y=1|x)$.

For the second component of the softmax function we would follow the same steps but have $a = p(y=1|x) - p(y=0|x)$.

Problem 4

To make the data linearly separable we could use a function like:

$$\phi(x_1, x_2) = (a_1, a_2) = (x_1, x_1 x_2) \tag{6}$$

This way class 'dot' will be above the line $y = 0$ since all the values in the second component of $\phi(x_1, x_2)$ are going to be positive, and class 'cross' will be below that line.