Machine Learning Worksheet Solution 5

Julius Jankowski

1 Linear separability

Problem 1:

A point on or inside the convex hull of a set of points $\mathbf{x}_i$ ($\mathbf{y}_j$ respectively) is given by:

$$\mathbf{x} = \sum_i \alpha_i \cdot \mathbf{x}_i \tag{1.1}$$

with $\sum_i \alpha_i = \sum_j \beta_j = 1$. An intersection means that there is a sum for which $\mathbf{x} = \mathbf{y}$:

$$\sum_i \alpha_i \cdot \mathbf{x}_i = \sum_j \beta_j \cdot \mathbf{y}_j \tag{1.2}$$

Without changing the statement of the equation, we can multiply both sides by $\mathbf{w}^T$ and subsequently add w_0 to both sides:

$$\begin{aligned}
\mathbf{w}^T(\sum_i \alpha_i \cdot \mathbf{x}_i) + w_0 &= \mathbf{w}^T(\sum_j \beta_j \cdot \mathbf{y}_j) + w_0 &(1.3)\\
\implies \sum_i \alpha_i \cdot \mathbf{w}^T\mathbf{x}_i + \sum_i \alpha_i \cdot w_0 &= \sum_j \beta_j \cdot \mathbf{w}^T\mathbf{y}_j + \sum_j \beta_j \cdot w_0 &(1.4)\\
\implies \sum_i \alpha_i \cdot (\mathbf{w}^T\mathbf{x}_i + w_0) &= \sum_j \beta_j \cdot (\mathbf{w}^T\mathbf{y}_j + w_0) &(1.5)
\end{aligned}$$

If we assume that the two sets $\mathbf{X}$ and $\mathbf{Y}$ are linearly seperable, we can find a vector $\mathbf{w}$ and a bias w_0 such that the resulting hyperplane splits the two sets. Consequently, $\mathbf{w}^T\mathbf{x}_i + w_0 > 0$ and $\mathbf{w}^T\mathbf{y}_j + w_0 < 0$ for all i and j. Since all α and β are supposed to be positive or zero, the intersection equation can not be satisfied.

Problem 2:

The logistic regression model estimates the real valued output by:

$$\hat{y} = \sigma(\mathbf{w}^T\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T\mathbf{x}}} \tag{1.6}$$

Consequently, the maximum likelihood error function looks like this:

$$E_{ML} = \frac{\beta}{2} \cdot \sum_{i=1}^{N}(y_i - \sigma(\mathbf{w}^T\mathbf{x}))^2 - \frac{N}{2} \cdot ln(\frac{\beta}{2\pi}) \tag{1.7}$$

In addition to that, we can say that a sample from class 1 has two properties: First, $\mathbf{w}^T\mathbf{x}_i > 0$ and second $y_i = 1$; And similarly for a sample from class 2: $\mathbf{w}^T\mathbf{z}_j < 0$ and $y_j = 0$. If we now let the norm of $\mathbf{w}$ grow to infinity, the logistic estimation is identical to the binary label:

$$\lim_{|\mathbf{w}|\to\infty} \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x_i}}} = 1 = y_i \tag{1.8}$$

$$\lim_{|\mathbf{w}|\to\infty} \frac{1}{1+e^{-\mathbf{w}^T\mathbf{z_j}}} = 0 = y_j \tag{1.9}$$

Hence, the logistic regression exactly hits the true values which of course minimizes the error function. We can prevent $|\mathbf{w}|$ growing to infinitiy by optimizing the maximum a posteriori error function instead of maximum likelihood.

Problem 3:

It's the well-known XOR-Problem, which can be solved by using:

$$\Phi(x1, x2) = x1 \cdot x2 \tag{1.10}$$

2 Basis functions

Problem 4:

Since our data $\mathbf{x} \in \mathbb{R}^2$ is 2-dimensional, the parameter vector is 2-dimensional as well $\mathbf{w} \in \mathbb{R}^2$. The bias is always one-dimensional $w_0 \in \mathbb{R}$. Consequently, we have 3 free parameters describing the boundary between the data sets. In contrast to that, we only have two linearly independent equations which have to be satisfied by the parameters:

$$\mathbf{w}^T \cdot \begin{pmatrix} 2 \\ 0 \end{pmatrix} + w_0 = 0 \tag{2.1}$$

$$\mathbf{w}^T \cdot \begin{pmatrix} 0 \\ 5 \end{pmatrix} + w_0 = 0 \tag{2.2}$$

Thus we can for example express the parameters w_1 and w_2 by w_0:

$$w_1 = -\frac{1}{2} w_0 \tag{2.3}$$

$$w_2 = -\frac{1}{5} w_0 \tag{2.4}$$

A valid set of parameters is: $w_0 = 1$, $w_1 = -\frac{1}{2}$ and $w_2 = -\frac{1}{5}$.