Machine Learnig Homework 7

Problem 1

If the data set is linearly separable, any decision boundary separating the two classes will have the property

$$w^T\phi(x^n) = \begin{cases} \geq 0 & \text{if } z^n = 1, \\ < 0 & otherwise. \end{cases} \tag{1}$$

Moreover, the log-likelihood

$$L(w) = \log p(\{z^i\}|w, \{\phi(x^i)\}) \tag{2}$$
$$= \sum_{n=1}^{N} \left[z^n \log \sigma(w^T\phi(x^n)) + (1 - z^n)(\log(1 - \sigma(w^T\phi(x^n)))) \right] \tag{3}$$

will be maximized when $y^n = \sigma(w^T\phi(x^n)) = z^n \quad \forall n$. This will be the case when the sigmoid function is saturated, which occurs when its argument, $w^T\phi$, goes to $\pm \inf$, i.e., when the magnitude of w goes to infinity.

Problem 2

The basis function

$$\phi(x_1, x_2) = x_1 x_2 \tag{4}$$

enables us to linearly seperate the crosses from the circles. The hyperplane would go through the origin of the coordinate system.

Problem 5

Let $\mathbf{w}^0 = \mathbf{0}$ be an initial set of weights and $b^0 = 0$ an initial bias. Let $z_k^i \in \{-1, 1\}$ denote the classification result of the i^{th} training sample $\mathbf{x}^i$, $i = 1, \ldots, N$ using weights $\mathbf{w}^k$ and bias b^k. The training algorithm iterates as long as there is at least one misclassification and computes:

$$\begin{aligned} \mathbf{w}^{k+1} &= \mathbf{w}^k + \sum_{t^i \neq z_k^i} t^i \mathbf{x}^i \\ b^{k+1} &= b^k + \sum_{t^i \neq z_k^i} t^i \end{aligned} \tag{5}$$

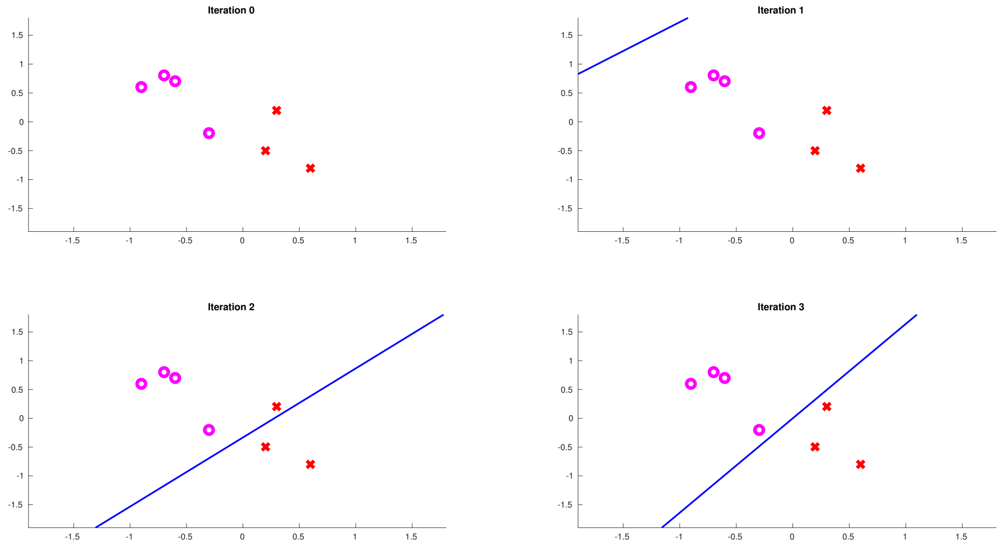

Figure 1: Decision boundary after 0-3 iterations.

Problem 6

Figure 1 shows the decision boundary after 0-3 training iterations. After the third iteration it does not change anymore; The norm of the weights $||\mathbf{w}^k||$ grows to infinity, however.

Problem 7

Let $z_k^i \in \{-1, 1\}$ denote the classification result of the i^{th} training sample $\mathbf{x}^i$, $i = 1, \dots, N$. Since $\mathbf{x}^i$ are linearly separable, it holds that $\mathbf{x}^i > 0 \implies t^i \widetilde{\mathbf{w}}^T \mathbf{x}^i > 0$. If we define $\gamma = \min_{\mathbf{x}^i} t^i \widetilde{\mathbf{w}}^T \mathbf{x}^i$ then:

$$\widetilde{\mathbf{w}}^T \mathbf{w}^k = \widetilde{\mathbf{w}}^T \mathbf{w}^{k-1} + \sum_{t^i \neq z_{k-1}^i} t^i \widetilde{\mathbf{w}}^T \mathbf{x}^i \geq \widetilde{\mathbf{w}}^T \mathbf{w}^{k-1} + \gamma \geq \widetilde{\mathbf{w}}^T \mathbf{w}^0 + k\gamma = k\gamma \tag{6}$$

Problem 8

From equation 5 and since $||\mathbf{x}^i|| < R$ and the number of misclassified samples at each iteration is at most N we have

$$\begin{aligned}||\mathbf{w}^k||^2 &\leq ||\mathbf{w}^{k-1}||^2 + \sum_{t^i \neq z^i_{k-1}} ||\mathbf{x}^i||^2 \\ &\leq ||\mathbf{w}^{k-1}||^2 + \sum_{t^i \neq z^i_{k-1}} R^2 \\ &\leq ||\mathbf{w}^{k-1}||^2 + NR^2 \\ &\leq kNR^2\end{aligned} \tag{7}$$

Problem 9

From problems 7 and 8 we know that $||\widetilde{\mathbf{w}}^T\mathbf{w}^k|| \geq k\gamma$ and $||\mathbf{w}^k||^2 < kR^2$. Therefore:

$$\frac{k^2\gamma^2}{kR^2} = \frac{k\gamma^2}{R^2} \leq \frac{\left(\widetilde{\mathbf{w}}^T\mathbf{w}^k\right)^2}{||\mathbf{w}^k||^2} \tag{8}$$

Since $\left(\widetilde{\mathbf{w}}^T\mathbf{w}^k\right)^2 = \sum_i \widetilde{\mathbf{w}}_i^2(\mathbf{w}^k)_i^2 = \sum_i \widetilde{\mathbf{w}}_i^2 \sum_i(\mathbf{w}^k)_i^2 = ||\widetilde{\mathbf{w}}||^2||\mathbf{w}^k||^2$ and

$$\frac{k\gamma^2}{R^2} \leq |||\widetilde{\mathbf{w}}||^2 \tag{9}$$

$$k \leq \frac{||\widetilde{R^2\mathbf{w}}||^2}{\gamma^2} \tag{10}$$