Cognitive Algorithms - Exercise Sheet 1

Solutions for * Exercise

Department of Machine Learning - TU Berlin

Task 3 - Convergence of the perceptron

1. Because the data set is linearly seperable, there exists a w_{sep}, such that for all x_i, $w_{sep}^T x_i \geq \xi$ for some $\xi \geq 0$. Let x_j be the data point with the largest squared norm $\|x_j\|^2$.

$$\begin{aligned}
&w_{sep}^T x_i y_i \geq \xi \\
&\qquad\Leftrightarrow \frac{1}{\xi} w_{sep}^T x_i y_i \geq 1 \\
&\qquad\Leftrightarrow [w'_{sep}]^T x_i y_i \geq 1 \\
&\qquad\Leftrightarrow [w''_{sep}]^T x_i y_i \geq \|x_j\|^2 \geq \|x_i\|^2
\end{aligned}$$

2. Now let $\mathbf{w}^{\text{sep}} = w''_{sep}$.

$$\begin{aligned}
\|\mathbf{w}^{\text{new}} - \mathbf{w}^{\text{sep}}\|^2 &= \|\mathbf{w}^{\text{old}} + \underbrace{\eta}_{=1} \mathbf{x}_m y_m - \mathbf{w}^{\text{sep}}\|^2 \\
&= (\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}} + \mathbf{x}_m y_m)^\top (\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}} + \mathbf{x}_m y_m) \\
&= (\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}})^\top (\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}) + 2(\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}})^\top \mathbf{x}_m y_m + (\mathbf{x}_m y_m)^\top \mathbf{x}_m y_m \\
&= \|\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}\|^2 + 2 \underbrace{(\mathbf{w}^{\text{old}})^\top \mathbf{x}_m y_m}_{\leq 0 \text{ since } \mathbf{x}_m \text{ misclassified}} - 2(\mathbf{w}^{\text{sep}})^\top \mathbf{x}_m y_m + \underbrace{y_m^2}_{=1} \|\mathbf{x}_m\|^2 \\
&\leq \|\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}\|^2 - 2(\mathbf{w}^{\text{sep}})^\top \mathbf{x}_m y_m + \|\mathbf{x}_m\|^2 \\
&\overset{3.1}{\leq} \|\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}\|^2 - 2\|\mathbf{x}_m\|^2 + \|\mathbf{x}_m\|^2 \\
&= \|\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}\|^2 - \|\mathbf{x}_m\|^2.
\end{aligned}$$