Problem set 5 Solution
Probabilistic classifiers 2

Excercise 1

1. Let $\overline{W}^T = (w_1, w_2, \ldots, w_d)$ and $\overline{P}^T = (p_1, p_2, \ldots, p_d)$ be points from R^d and let b be a number from R. Compute the distance from $\overline{P}$ to the hyperplane defined by the equation $\overline{W}^T\overline{X} + b = 0$.

2. Show that the distance is proportional to $|\overline{W}^T\overline{P} + b|$, that is the distance is equal to $c \cdot |\overline{W}^T\overline{P} + b|$ for some number c.

Solution 1

We use the fact that the vector $\overline{W}$ is perpendicular to the hyperplane $\overline{W}^T\overline{X}+b = 0$.

Let L be the line via the point $\overline{P}$ and parallel to the vector $\overline{W}$, that is $L = \{\overline{P} + t\overline{W} \mid t \in R\} \subseteq R^d$. Line L intersects the hyperplane $\overline{W}^T\overline{X} + b = 0$ when $\overline{W}^T(\overline{P} + t\overline{W}) + b = 0$, i.e. when

$$t = \frac{-(\overline{W}^T\overline{P} + b)}{||\overline{W}||^2}.$$

The distance between this point of intersection and the starting point $\overline{P}$ is

$$d = ||\overline{P} + t\overline{W} - \overline{P}|| = ||t\overline{W}|| = |t| \cdot ||\overline{W}|| = \frac{|\overline{W}^T\overline{P} + b|}{||\overline{W}||^2} \cdot ||\overline{W}|| = \frac{|\overline{W}^T\overline{P} + b|}{||\overline{W}||}.$$

The latter is proportional to $|\overline{W}^T\overline{P} + b|$.

Solution 2

In this solution we use the method of Lagrange multipliers. We want to find a point $\overline{X}$ in the hyperplane (meaning that $\overline{W}^T\overline{X} + b = 0$) that has the smallest distance to $\overline{P}$ (meaning that $||\overline{P} - \overline{X}||$ is minimized).

In other words, we want to solve the following optimisation problem:
$\min\limits_{\overline{X}\in R^d} (\overline{P}-\overline{X})^T(\overline{P}-\overline{X})$
subject to $\overline{W}^T\overline{X}+b=0$.
The corresponding Lagrangian is $\mathcal{L}(x_1,\ldots,x_d,\lambda)=(\overline{P}-\overline{X})^T(\overline{P}-\overline{X})-\lambda\cdot(\overline{W}^T\overline{X}+b)$.
The derivative of the Lagrangian is $2(\overline{P}-\overline{X})-\lambda\overline{W}$.
To solve $2(\overline{P}-\overline{X})-\lambda\overline{W}=0$ we first take the dot product of each of the sides of the equation with vector $\overline{W}$:

$$2\overline{W}^T(\overline{P}-\overline{X})-\lambda\cdot\overline{W}^T\overline{W}=0,$$

from which we derive $\lambda=\frac{2\overline{W}^T(\overline{P}-\overline{X})}{||\overline{W}||^2}$.
Similarly, we take the dot product with $(\overline{P}-\overline{X})$:

$$2(\overline{P}-\overline{X})^T(\overline{P}-\overline{X})-\lambda\cdot(\overline{P}-\overline{X})^T\overline{W}=0,$$

from which, using the computed value of λ, we derive

$$2(\overline{P}-\overline{X})^T(\overline{P}-\overline{X})=\frac{2\overline{W}^T(\overline{P}-\overline{X})}{||\overline{W}||^2}\cdot(\overline{P}-\overline{X})^TW$$

$$||\overline{P}-\overline{X}||^2=\frac{(\overline{W}^T(\overline{P}-\overline{X}))^2}{||\overline{W}||^2}$$

$$||\overline{P}-\overline{X}||^2=\frac{(\overline{W}^T\overline{P}+b-\overline{W}^T\overline{X}-b)^2}{||\overline{W}||^2}=\frac{(\overline{W}^T\overline{P}+b)^2}{||\overline{W}||^2}$$

$$||\overline{P}-\overline{X}||=\frac{|\overline{W}^T\overline{P}+b|}{||\overline{W}||}$$

Excercise 2

Show the following properties of the logistic sigmoid function $\sigma(x)=\frac{1}{1+e^{-x}}$.

1. $\sigma(x)=1-\sigma(-x)$
2. $\frac{\partial\sigma}{\partial x}=e^{-x}\sigma^2(x)$

Solution

1. $1-\sigma(-x)=1-\frac{1}{1+e^x}=\frac{1+e^x-1}{1+e^x}=\frac{e^x}{1+e^x}\cdot\frac{e^{-x}}{e^{-x}}=\frac{1}{e^{-x}+1}=\sigma(x)$
2. $\frac{\partial\sigma}{\partial x}=\frac{e^{-x}}{(1+e^{-x})^2}=e^{-x}\sigma^2(x)$