# Pattern Recognition Assignment#3

May 2, 2023

Q1. Suppose we have four normalized training samples under the two-category case: $\boldsymbol{y}_1 = (1,4,1)^\top, \boldsymbol{y}_2 = (1,2,2)^\top, \boldsymbol{y}_3 = (-1,0,-2)^\top, \boldsymbol{y}_4 = (-1,-1,-1)^\top$. The generalized linear discriminant function $g(\boldsymbol{y}) = \boldsymbol{a}^\top \boldsymbol{y}$ is adopted to learn from the training samples and the criterion function to be minimized is set as $J_p(\boldsymbol{a}) = \Sigma_{\boldsymbol{y}\in\gamma}(-\boldsymbol{a}^\top \boldsymbol{y})$.

(a) Given an initial model $\boldsymbol{a} = (0,0,0)^\top$, if the **fixed-increment single-sample correction algorithm** is utilized to minimize the criterion function, what is the final resulting discriminant function with fixed learning rate $\eta = 1$ ?

(b) Given an initial model $\boldsymbol{a} = (-3,-1,1)^\top$, if the **batch perceptron algorithm** is utilized to minimize the criterion function, what is the final resulting discriminant function with fixed learning rate $\eta = 0.5$ and threshold $\theta = 0.5$ ?

Q2. One of the simplest approaches for feature selection is the filter method, in which we assess individual features, independently of other features, according to some quality measure, and then select the $k$ features that achieve the highest score. Consider a linear regression problem with the squared loss. Let $\boldsymbol{v} = (v_{1,j}, \ldots, v_{m,j}) \in \mathbb{R}^m$ be a vector designating the values of the $j$-th feature on a training set of $m$ examples and let $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ be the values of the target on the same $m$ examples. The empirical squared loss of an empirical risk minimization linear predictor that uses only the $j$-th feature would be

$$\min_{a,b\,\in\,\mathbb{R}} \frac{1}{m}\|a\boldsymbol{v} + b - \boldsymbol{y}\|^2,$$

where the meaning of adding a scalar $b$ to a vector $\boldsymbol{v}$ is adding $b$ to all coordinates of $\boldsymbol{v}$, and multiplying a vector $\boldsymbol{v}$ by a scalar $a$ is multiplying all coordinates of $\boldsymbol{v}$ by $a$. To solve this problem, let $\bar{v} = \frac{1}{m}\sum_{i=1}^m v_i$ be the mean value of the feature and let $\bar{y} = \frac{1}{m}\sum_{i=1}^m y_i$ be the mean value of the target. It clearly holds that,

$$\min_{a,b\,\in\,\mathbb{R}} \frac{1}{m}\|a\boldsymbol{v} + b - \boldsymbol{y}\|^2 = \min_{a,b\,\in\,\mathbb{R}} \frac{1}{m}\|a(\boldsymbol{v} - \bar{v}) + b - (\boldsymbol{y} - \bar{y})\|^2. \tag{1}$$

Taking the derivative of the right-hand side of Equation (1) with respect to $b$ and comparing it to zero, we obtain that $b = 0$. Similarly, solving for $a$ (once we know that $b = 0$) yields $a = \langle \boldsymbol{v} - \bar{v}, \boldsymbol{y} - \bar{y}\rangle / \|\boldsymbol{v} - \bar{v}\|^2$. Plugging this value back into Equation (1), we can obtain that

$$\|\boldsymbol{y} - \bar{y}\|^2 - \frac{(\langle \boldsymbol{v} - \bar{v}, \boldsymbol{y} - \bar{y}\rangle)^2}{\|\boldsymbol{v} - \bar{v}\|^2}.$$

Ranking the features according to the minimal value of the loss is equivalent to ranking them according to the absolute value of the following score (where a higher score yields a better feature):

$$\frac{\langle \boldsymbol{v} - \bar{v}, \boldsymbol{y} - \bar{y} \rangle}{\|\boldsymbol{v} - \bar{v}\| \|\boldsymbol{y} - \bar{y}\|} = \frac{\frac{1}{m}\langle \boldsymbol{v} - \bar{v}, \boldsymbol{y} - \bar{y} \rangle}{\sqrt{\frac{1}{m}\|\boldsymbol{v} - \bar{v}\|^2}\sqrt{\frac{1}{m}\|\boldsymbol{y} - \bar{y}\|^2}}.$$

The preceding expression is known as Pearson's correlation coefficient. Prove the equality given in Equation (1).

**Hint:** Let $a^*$, $b^*$ be minimizers of the left-hand side. Find $a$, $b$ such that the loss value of the right-hand side is smaller than that of the left-hand side. Do the same for the other direction.

Q3. Given three samples $\boldsymbol{x}_1 = (2, 2, 3)^\top$, $\boldsymbol{x}_2 = (4, 0, 3)^\top$ and $\boldsymbol{x}_3 = (3, 1, -1)^\top$, please reduce the original 3-dimensional samples to 1-dimensional samples using **principal component analysis** (PCA).

Q4. Given two sample sets $D_1$ and $D_2$, whose categories are $\omega_1$ and $\omega_2$ respectively. $D_1$ includes three samples $\boldsymbol{x}_1 = (1, 3)^\top, \boldsymbol{x}_2 = (3, 7)^\top, \boldsymbol{x}_3 = (2, 2)^\top$, and $D_2$ includes two samples $\boldsymbol{x}_4 = (-1, -3)^\top, \boldsymbol{x}_5 = (-3, -7)^\top$. Please write the specific **linear discriminant analysis** (LDA) criterion function (i.e., generalized Rayleigh quotient) $J(\boldsymbol{w}) = \frac{\boldsymbol{w}^\top \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{S}_W \boldsymbol{w}}$ of these samples, where $\boldsymbol{S}_B$ is between-class scatter matrix and $\boldsymbol{S}_W$ is within-class scatter matrix.

**Hint:** Calculate $\boldsymbol{S}_B$ and $\boldsymbol{S}_W$.