Task 1 - Covariance matrix

1. Consider a data set with four data points: $\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}-1\\0\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}$. Compute the covariance matrix.

2. Consider N d-dimensional data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^d$ with mean zero, concatenated in a data matrix $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{d\times N}$. Their covariance can be computed as $S := \frac{1}{N}XX^T = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\top$. Our goal is to compute the variance of the data projected onto a vector $\mathbf{w} \in \mathbb{R}^d$.

   Since our data points have mean zero, the variance of the projected data is defined as $\frac{1}{N}\sum_{i=1}^N(\mathbf{w}^T\mathbf{x}_i)^2$. Show:

$$\frac{1}{N}\sum_{i=1}^N(\mathbf{w}^T\mathbf{x}_i)^2 = \mathbf{w}^T S\mathbf{w}$$

Task 2 - Bias term of the Nearest Centroid Classifier

Remember the Nearest Centroid/ Prototype classifier. For it's classification boundary $\mathbf{w}^\top\mathbf{x} - \beta = 0$ we obtained

$$\mathbf{w} = \mathbf{w}_o - \mathbf{w}_\Delta$$

$$\beta = 1/2 \cdot (\mathbf{w}_o^\top\mathbf{w}_o - \mathbf{w}_\Delta^\top\mathbf{w}_\Delta)$$

where $\mathbf{w}_o$ and $\mathbf{w}_\Delta$ denote the respective class means. Show that:

$$\beta = 1/2 \cdot \mathbf{w}^\top(\mathbf{w}_o + \mathbf{w}_\Delta)$$

Task 3 - Whitening

Whitening is a linear transformation of a given data set into a new data set whose covariance matrix is the identiy matrix (i.e. the features are uncorrelated and all have variance 1). For many algorithms, whitening is a useful preprocessing step.

Suppose we have N d-dimensional data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^d$ with mean zero. It's $d \times d$ covariance matrix can then be computed as $S = 1/N\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\top$. Let $S = U\Lambda U^T$ be the eigendecomposition of S, where Λ is a diagonal matrix and U is an orthogonal matrix (i.e. $UU^\top = U^\top U = I$). The whitening transformation is then given as

$$\mathbf{z}_i := \Lambda^{-\frac{1}{2}}U^\top\mathbf{x}_i$$

(where $\Lambda^{\frac{1}{2}}$ is the matrix square root of Λ, i.e. $\Lambda^{\frac{1}{2}} \cdot \Lambda^{\frac{1}{2}} = \Lambda$ and $\Lambda^{-\frac{1}{2}} \cdot \Lambda^{\frac{1}{2}} = I$)
Show that the covariance of the whitened data $\mathbf{z}_1, \ldots, \mathbf{z}_N$ is the identity, i.e. show that

$$\frac{1}{N}\sum_{i=1}^N \mathbf{z}_i\mathbf{z}_i^\top = I$$

Task 4 - Maximum Likelihood Estimator

In statistics, the Bayesian estimator minimizes the posterior expected value of a loss function. In the case of binary classification, this means that a Bayes estimator considers the class with a higher posteriori probability, e.g. if

$$p(\Delta|x) > p(\circ|x) \tag{1}$$

a Bayes estimator would assign class Δ.
For LDA, we assume Gaussian distributed data, s.t.

$$p(x|\Delta) = \frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|S_\Delta|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{w}_\Delta)^\top S_\Delta^{-1}(\mathbf{x}-\mathbf{w}_\Delta)}$$

Compute the Bayes classifier that fulfills (1), remember that

$$\begin{aligned} p(\Delta|x) &> p(\circ|x) \\ \Leftrightarrow \frac{p(\Delta|x)}{p(\circ|x)} &> 1 \\ \Leftrightarrow \log\left(\frac{p(\Delta|x)}{p(\circ|x)}\right) &> 0. \end{aligned}$$

Check slides 23 and 24 of Lecture 3 for further information.

Solutions

Task1

1.

$$X = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned} S =& \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^\top = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^\top \\ =& \frac{1}{N}XX^\top = \frac{1}{4}\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \end{aligned}$$

2.

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}^\top\mathbf{x}_i)^2 &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}^\top\mathbf{x}_i)(\mathbf{w}^\top\mathbf{x}_i)^\top \\ &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}^\top\mathbf{x}_i)(\mathbf{x}_i^\top\mathbf{w}) \\ &= \mathbf{w}^\top\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^\top\mathbf{w} \\ &= \mathbf{w}^\top S\mathbf{w} \end{aligned}$$

Task2

$$\begin{aligned} \beta &= 1/2\cdot(\mathbf{w}_o^\top\mathbf{w}_o - \mathbf{w}_\Delta^\top\mathbf{w}_\Delta) \\ &= 1/2\cdot(\mathbf{w}_o - \mathbf{w}_\Delta)^\top(\mathbf{w}_o + \mathbf{w}_\Delta) \\ &= 1/2\cdot\mathbf{w}^\top(\mathbf{w}_o + \mathbf{w}_\Delta) \end{aligned}$$

Task3

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i^\top &= \frac{1}{N}\sum_{i=1}^{N}(\Lambda^{-\frac{1}{2}}U^\top\mathbf{x}_i)(\Lambda^{-\frac{1}{2}}U^\top\mathbf{x}_i)^\top && \text{as } U \text{ and } \Lambda \text{ do not depend on } i \\ &= \Lambda^{-\frac{1}{2}}U^\top\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^\top U\Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}}U^\top SU\Lambda^{-\frac{1}{2}} && \text{eigendecomposition of } S \\ &= \Lambda^{-\frac{1}{2}}U^\top U\Lambda U^\top U\Lambda^{-\frac{1}{2}} && U \text{ is orthogonal} \\ &= \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}}\Lambda^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}\Lambda^{\frac{1}{2}} = I \end{aligned}$$