Applied Machine Learning

HW2-Written Exercises

1. HTF Exercise 4.1 (From Generalized to Standard Eigenvalue Problem)

Ex. 4.1 Show how to solve the generalized eigenvalue problem $\max a^T \mathbf{B} a$ subject to $a^T \mathbf{W} a = 1$ by transforming to a standard eigenvalue problem.

By Lagrange multiplies, we can define $L(a) = a^T B a + \lambda(a^T W a - 1)$

$$\frac{\partial L(a)}{\partial a} = 2Ba + \lambda(2Wa) = 0$$

$$\Rightarrow Ba + \lambda Wa = 0$$

$$\Rightarrow W^{-1}Ba = \lambda a \text{, which is a standard eigenvalue problem.}$$

2. HTF Exercise 4.2 (Correspondence between LDA and classification by linear least squares.)

Ex. 4.2 Suppose we have features $x \in \mathbb{R}^p$, a two-class response, with class sizes N_1, N_2, and the target coded as $-N/N_1, N/N_2$.

(a) Show that the LDA rule classifies to class 2 if

$$x^T \hat{\mathbf{\Sigma}}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2}(\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\mathbf{\Sigma}}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) - \log(N_2/N_1),$$

and class 1 otherwise.

(b) Consider minimization of the least squares criterion

$$\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T \beta)^2. \tag{4.55}$$

Show that the solution $\hat{\beta}$ satisfies

$$\left[(N-2)\hat{\mathbf{\Sigma}} + N\hat{\mathbf{\Sigma}}_B\right]\beta = N(\hat{\mu}_2 - \hat{\mu}_1) \tag{4.56}$$

(after simplification), where $\hat{\Sigma}_B = \frac{N_1 N_2}{N^2}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$.

(c) Hence show that $\hat{\Sigma}_B \beta$ is in the direction $(\hat{\mu}_2 - \hat{\mu}_1)$ and thus

$$\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1). \tag{4.57}$$

Therefore the least-squares regression coefficient is identical to the LDA coefficient, up to a scalar multiple.

(d) Show that this result holds for any (distinct) coding of the two classes.

(e) Find the solution $\hat{\beta}_0$ (up to the same scalar multiple as in (c), and hence the predicted value $\hat{f}(x) = \hat{\beta}_0 + x^T\hat{\beta}$. Consider the following rule: classify to class 2 if $\hat{f}(x) > 0$ and class 1 otherwise. Show this is not the same as the LDA rule unless the classes have equal numbers of observations.

(Fisher, 1936; Ripley, 1996)

(a) The LDA rule classifies to class 2 if $\delta_1(x) < \delta_2(x)$

since $\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k$

$\pi_k = \frac{N_k}{N}$

if $\delta_1(x) < \delta_2(x)$, $x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \log\frac{N_2}{N} > x^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\frac{N_1}{N}$

$$\Rightarrow \quad x^T\Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2}(\mu_2 + \mu_1)^T\Sigma^{-1}(\mu_2 - \mu_1) - \log\frac{N_2}{N_1}$$

(b) To minimize the least squares criterion:

$\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2$, $(\hat{\beta}_0, \hat{\beta})'$ should have: $X^TX\begin{bmatrix}\beta_0\\ \beta\end{bmatrix} = X^Ty$.

$$X^TX = \begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_{N_1+N_2}\end{bmatrix}\begin{bmatrix}1 & x_1^T\\ 2 & x_2^T\\ \vdots & \\ 1 & x_{N_1+N_2}^T\end{bmatrix} = \begin{bmatrix}N & \sum_{i=1}^{N} x_i^T\\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i x_i^T\end{bmatrix}$$

Since the targets codes as $-N/N_1$, N/N_2, $N = N_1 + N_2$

$$X^Ty = \begin{bmatrix}1 & 2 & \cdots & 1\\ x_1 & x_2 & \cdots & x_{N_1+N_2}\end{bmatrix}\begin{bmatrix}-N/N_1\\ \vdots\\ -N/N_1\\ N/N_2\\ \vdots\\ N/N_2\end{bmatrix}$$

$$= \begin{bmatrix}N_1(-\frac{N}{N_1}) + N_2(\frac{N}{N_2})\\ (\sum_{i=1}^{N_1} x_i)(-\frac{N}{N_1}) + (\sum_{i=N_1+1}^{N} x_i)(\frac{N}{N_2})\end{bmatrix}$$

$$= \begin{bmatrix}0\\ -N\mu_1 + N\mu_2\end{bmatrix}$$

The pooled covariance matrix $\hat{\Sigma} = \frac{1}{N-k}\sum_{k=1}^{k}\sum_{i:g_i=k}(x_i-\mu_k)(x_i-\mu_k)^T$

Since $k=2$ in this case,

$$\hat{\Sigma} = \frac{1}{N-2}\left[\sum_{i:g_i=1} x_i x_i^T - N_1\mu_1\mu_1^T + \sum_{i:g_i=1} x_i x_i^T - N_2\mu_2\mu_2^T\right]$$

$$\sum_{i=1}^{N} x_i x_i^T = (N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T$$

$$X^T X \begin{bmatrix}\beta_0\\ \beta\end{bmatrix} = X^T y$$

$$\Rightarrow \begin{bmatrix} N & N_1\mu_1^T + N_2\mu_2^T \\ N_1\mu_1 + N_2\mu_2 & (N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta\end{bmatrix} = \begin{bmatrix}0\\ -N\mu_1 + N\mu_2\end{bmatrix}$$

$$N\beta_0 + (N_1\mu_1^T + N_2\mu_2^T)\beta = 0$$

$$\Rightarrow \beta_0 = \left(-\frac{N_1}{N}\mu_1^T - \frac{N_2}{N}\mu_2^T\right)\beta$$

$$\Rightarrow (N_1\mu_1 + N_2\mu_2)\left(-\frac{N_1}{N}\mu_1^T - \frac{N_2}{N}\mu_2^T\right)\beta + \left[(N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T\right]\beta = N(\mu_2-\mu_1)$$

denote $\hat{\Sigma}_B = (\mu_2-\mu_1)(\mu_2-\mu_1)^T$

$$\Rightarrow \left[(N-2)\hat{\Sigma} + \frac{N_1N_2}{N}\hat{\Sigma}_B\right]\beta = N(\mu_2-\mu_1)$$

(c) $\hat{\Sigma}_B\beta = (\mu_2-\mu_1)(\mu_2-\mu_1)^T\beta = \lambda(\hat{\mu}_2-\hat{\mu}_1)$

the direction of $\hat{\Sigma}_B\beta$ is the same as the direction of $\mu_2-\mu_1$

$$\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)$$

(d) The above proof holds for any arbitrary and distinct $\frac{N_1}{N}, \frac{N_2}{N}$.

(e) $\hat{\beta}_0 = -\frac{1}{N}(N_1\hat{\mu}_1^T + N_2\hat{\mu}_2^T)\hat{\beta}$

$$\hat{\beta}_0 + \hat{\beta}^T = \frac{1}{N}(Nx^T - N_1\hat{\mu}_1^T - N_2\hat{\mu}_2^T)\lambda\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)$$

classify by considering if $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}^T x > 0$

$$x^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1) > \frac{1}{N}(N_1\hat{\mu}_1^T + N_2\hat{\mu}_2^T)\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)$$

which is different from LDA

3.RLU Exercise 11.3.1 (SVD of Rank Deficient Matrix)

Exercise 11.3.1: In Fig. 11.11 is a matrix M. It has rank 2, as you can see by observing that the first column plus the third column minus twice the second column equals **0**.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$$

Figure 11.11: Matrix M for Exercise 11.3.1

(a) Compute the matrices $M^{\mathrm{T}}M$ and MM^{T}.

! (b) Find the eigenvalues for your matrices of part (a).

(c) Find the eigenvectors for the matrices of part (a).

(d) Find the SVD for the original matrix M from parts (b) and (c). Note that there are only two nonzero eigenvalues, so your matrix Σ should have only two singular values, while U and V have only two columns.

(e) Set your smaller singular value to 0 and compute the one-dimensional approximation to the matrix M from Fig. 11.11.

(f) How much of the energy of the original singular values is retained by the one-dimensional approximation?

(a) M^TM

$$= \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 36 & 37 & 38 \\ 37 & 49 & 61 \\ 38 & 61 & 84 \end{bmatrix}$$

MM^T

$$= \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 14 & 26 & 22 & 16 & 22 \\ 26 & 50 & 46 & 28 & 40 \\ 22 & 46 & 50 & 20 & 32 \\ 16 & 28 & 20 & 20 & 26 \\ 22 & 40 & 32 & 26 & 35 \end{bmatrix}$$

(b) For M^TM,

(c) $$\begin{vmatrix} 36-\lambda & 37 & 38 \\ 37 & 49-\lambda & 61 \\ 38 & 61 & 84-\lambda \end{vmatrix} = -\lambda^3 + 169\lambda^2 - 2370\lambda$$

$$\Rightarrow \lambda_1 = 0 \quad v_1 = \begin{bmatrix} -0.4 \\ 0.8 \\ -0.4 \end{bmatrix}$$

$$\lambda_2 = \frac{\sqrt{19081} + 169}{2} \approx 153.6 \quad v_2 = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.7 \end{bmatrix}$$

$$\lambda_3 = \frac{-\sqrt{19081} + 169}{2} \approx 15.4 \quad v_3 = \begin{bmatrix} 0.8 \\ 0.1 \\ -0.6 \end{bmatrix}$$

For MM^T

$$\begin{vmatrix} 16-\lambda & 26 & 22 & 16 & 22 \\ 26 & 50-\lambda & 46 & 28 & 40 \\ 22 & 46 & 50-\lambda & 20 & 32 \\ 16 & 28 & 20 & 20-\lambda & 26 \\ 22 & 40 & 32 & 26 & 35-\lambda \end{vmatrix} =$$

$= -\lambda^5 + 169\lambda^4 - 23760\lambda^3$

$\lambda_1 = \lambda_2 = \lambda_3 = 0$, $\lambda_4 = \frac{-\sqrt{19081} + 169}{2} \approx 15.4$

$\lambda_5 = \frac{\sqrt{19081} + 169}{2} \approx 153.6$

eigen vectors:

$$\begin{bmatrix} -0.2 \\ -0.5 \\ 0.2 \\ -0.3 \\ 0.8 \end{bmatrix}, \begin{bmatrix} 0.9 \\ -0.3 \\ 0.04 \\ -0.2 \\ -0.1 \end{bmatrix}, \begin{bmatrix} 0.1 \\ 0.5 \\ -0.4 \\ -0.7 \\ 0.2 \end{bmatrix}, \begin{bmatrix} 0.2 \\ -0.03 \\ -0.7 \\ 0.5 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 0.3 \\ 0.6 \\ 0.5 \\ 0.3 \\ 0.5 \end{bmatrix}$$

(d) U: eigenvector of MM^T

$$U = \begin{bmatrix} 0.3 & 0.2 \\ 0.6 & -0.03 \\ 0.5 & -0.7 \\ 0.3 & 0.5 \\ 0.5 & 0.4 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{153.6} & 0 \\ 0 & \sqrt{15.4} \end{bmatrix} = \begin{bmatrix} 12.4 & 0 \\ 0 & 3.9 \end{bmatrix}$$

V: eigenvector of M^TM

$$V^T = \begin{bmatrix} 0.4 & 0.6 & 0.7 \\ -0.8 & -0.1 & 0.6 \end{bmatrix}$$

$M = U\Sigma V^T$

(e) $$\begin{bmatrix} 0.3 & 0.2 \\ 0.6 & -0.03 \\ 0.5 & -0.7 \\ 0.3 & 0.5 \\ 0.5 & 0.4 \end{bmatrix} \begin{bmatrix} 12.4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.4 & 0.6 & 0.7 \\ -0.8 & -0.1 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 1.5 & 2.1 & 2.5 \\ 2.9 & 4.0 & 5.1 \\ 2.6 & 3.7 & 4.6 \\ 1.6 & 2.3 & 2.9 \\ 2.3 & 3.2 & 4.1 \end{bmatrix}$$

(f) $\frac{153.6}{153.6 + 15.4} \approx 90.9\%$