Formulas from 'Dimensionality Reduction'

1 Principal Component Analysis (PCA)

1.1 Reconstruction error minimization criterion (Objective function)

$$P^* = \arg\min_P \frac{1}{K}\sum_{i=1}^{K} \|x_i - \hat{x}_i\|^2 = \arg\min_P \frac{1}{K}\sum_{i=1}^{K} \left\|x_i - PP^T x_i\right\|^2$$

where K is the number of samples

1.2 PCA solution (Centered Data)

Compute eigen-decomposition of covariance matrix:

$$\frac{1}{K}\sum_{i=1}^{K} x_i x_i^T = U\Sigma U^T \tag{1}$$

Select m eigenvectors of the matrix U corresponding to the m highest eigenvalues in matrix Σ:

$$P^* = [u_1, ..., u_m] \tag{2}$$

The projection of $\mathbf{x}$ over the subspace is given by

$$\mathbf{y} = P^T\mathbf{x}$$

1.3 PCA solution (Non-Centered Data)

- Compute sample mean $\bar{x}$
- Center data: $z_i = x_i - \bar{x}$
- Compute the sample covariance matrix

$$C = \frac{1}{K}\sum_i (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{K}\sum_i z_i z_i^T$$

- Compute the eigen–decomposition of $C = U\Sigma U^T$

- Project the data in the subspace spanned by the m columns of U corresponding to the m highest eigenvalues (matrix P):

$$y_i = P^T z_i = P^T (x_i - \bar{x})$$

- Reconstruction requires inverting the process: $\hat{x}_i = \bar{P} y_i + \bar{x}$

1.4 Variance explained criterion for choosing m

$$\min_m \quad \sum_{i=1}^{m} \sigma_i \bigg/ \sum_{i=1}^{n} \sigma_i \geq t \tag{3}$$

2 Linear Discriminant Analysis (LDA)

2.1 Between-class and within-class scatter matrices

$$S_B = \frac{1}{N} \sum_{c=1}^{K} n_c (\mu_c - \mu)(\mu_c - \mu)^T \tag{4}$$

$$S_W = \frac{1}{N} \sum_{c=1}^{K} \sum_{i=1}^{n_c} (x_{c,i} - \mu_c)(x_{c,i} - \mu_c)^T \tag{5}$$

μ is the dataset mean. μ_c is the class mean.

2.2 Optimization criterion (Objective function)

Maximize the following function:

$$\mathcal{L}(w) = \frac{w^T S_B w}{w^T S_W w}$$

We can observe that

The optimal solution is an eigenvector of $S_W^{-1} S_B$.

The eigenvalue corresponding to solution w is $\lambda(w) = \mathcal{L}(w)$, i.e. the value of the ratio we want to maximize

2.3 LDA for Binary Classification

$$w \propto S_W^{-1} (\mu_2 - \mu_1) \tag{6}$$

$$C(x_t) = \begin{cases} C_1, & \text{if } w^T x_t < t \\ C_2, & \text{if } w^T x_t \geq t \end{cases} \tag{7}$$

2.4 LDA for dimensionality reduction

We can express the projected (in the new feature space) between and within class covariance matrices as

$$\hat{S}_B = W^T S_B W$$
$$\hat{S}_W = W^T S_W W$$

Different criteria can be used to generalize the 1-dimensional case.

A common one looks for the maximizer of

$$\mathcal{L} = \mathrm{Tr}\left(\hat{S}_W^{-1}\hat{S}_B\right)$$

It can be shown that the solution is given by the m (right) eigenvectors corresponding to the m largest eigenvalues of $S_W^{-1}S_B$.

We can also compute the solution by solving the generalized eigenvalue problem $S_B w = \lambda S_W w$.

Therefore, LDA allows estimating at most $C-1$ directions.