

Tutorial 3: Dimensionality Reduction

Rui Zhao

rzhao@ee.cuhk.edu.hk

Feb. 6, 2014

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

1. Lagrange Optimization

Constrained optimization problem:

$$\max_{\mathbf{x}} f(\mathbf{x}), \quad (1)$$

$$s.t. \quad g(\mathbf{x}) = 0. \quad (2)$$

The solution can often be found by Lagrangian method. The Lagrangian is defined as:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x}). \quad (3)$$

1. Lagrange Optimization

Lagrangian Sufficiency Theorem: Suppose there exist $\mathbf{x}^* \in \mathbf{X}$ and λ^* , such that \mathbf{x}^* maximize $L(\mathbf{x}, \lambda^*)$ over all $\mathbf{x} \in \mathbf{X}$, and $g(\mathbf{x}^*) = 0$. Then \mathbf{x}^* solves the optimization problem.

Proof.

$$\max_{\mathbf{x} \in \mathbf{X}, g(\mathbf{x})=0} f(\mathbf{x}) \quad (4)$$

$$= \max_{\mathbf{x} \in \mathbf{X}, g(\mathbf{x})=0} [f(\mathbf{x}) + \lambda^* g(\mathbf{x})] \quad (5)$$

$$\leq \max_{\mathbf{x} \in \mathbf{X}} [f(\mathbf{x}) + \lambda^* g(\mathbf{x})] \quad (6)$$

$$= [f(\mathbf{x}^*) + \lambda^* g(\mathbf{x}^*)] \quad (7)$$

$$= f(\mathbf{x}^*) \quad (8)$$

1. Lagrange Optimization

Solve Lagrange Optimization: solve the unconstrained problem by taking the derivative w.r.t. \mathbf{x} and λ :

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 0 \quad (9)$$

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} = 0 \quad (10)$$

$$\text{or} \quad (11)$$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \lambda \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = 0 \quad (12)$$

$$g(\mathbf{x}) = 0 \quad (13)$$

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

2. Generalized Rayleigh Quotient

Fisher Criterion

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}} \quad (14)$$

$J(\mathbf{w})$ is the generalized Rayleigh quotient. A vector \mathbf{w} that maximizes $J(\cdot)$ must satisfy

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \quad (15)$$

for some constant λ .

2. Generalized Rayleigh Quotient

Maximizing $J(\mathbf{w})$ is equivalent to

$$\max_{\mathbf{w}} \quad \mathbf{w}^t \mathbf{S}_B \mathbf{w} \quad (16)$$

$$s.t. \quad \mathbf{w}^t \mathbf{S}_W \mathbf{w} = K \quad (17)$$

which can be solved using Lagrange multipliers.

2. Generalized Rayleigh Quotient

Define the Lagrangian:

$$L = \mathbf{w}^t \mathbf{S}_B \mathbf{w} - \lambda(\mathbf{w}^t \mathbf{S}_W \mathbf{w} - K) \quad (18)$$

Maximize with respect to \mathbf{w} :

$$\nabla_{\mathbf{w}} L = 2(\mathbf{S}_B - \lambda \mathbf{S}_W) \mathbf{w} = 0 \quad (19)$$

To obtain the solution:

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \quad (20)$$

2. Generalized Rayleigh Quotient

Generalized eigenvalue problem:

$$\mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w} \quad (21)$$

- $\mathbf{S}_B \mathbf{w} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t \mathbf{w} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)a$ is always the direction $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. Thus $\mathbf{w}^* = \mathbf{S}_W^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$.
- For an optimal \mathbf{w}^* , $\mathbf{w}^{*t} \mathbf{S}_B \mathbf{w}^* = \lambda \mathbf{w}^{*t} \mathbf{S}_W \mathbf{w}^* = \lambda K$, $J(\mathbf{w})$ is maximized by the largest eigenvalue.
 - \mathbf{S}_W is invertible. \mathbf{w}^* is the eigenvector corresponding to the largest eigenvalue of $\mathbf{S}_W^{-1} \mathbf{S}_B$.
 - \mathbf{S}_W is not invertible. \mathbf{w}^* is the eigenvector corresponding to the largest eigenvalue of $[\mathbf{S}_W^{-1} + \alpha I]^{-1} \mathbf{S}_B$, which is equivalent to a new regularized problem:

$$\max_{\mathbf{w}} \mathbf{w}^t \mathbf{S}_B \mathbf{w}, \text{ s.t. } \mathbf{w}^t \mathbf{S}_W \mathbf{w} = K, \|\mathbf{w}\| = L$$

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

Outline

- 1 Lagrange Optimization
- 2 Generalized Rayleigh Quotient
- 3 Exercises

3. Exercises

Chapter 4. Problem 35

Consider two normal distribution with arbitrary but equal covariances. Prove that the Fisher linear discriminant, for suitable threshold, can be derived from the negative of the log-likelihood ratio.

3. Exercises

Prove.

- Recall normal distribution.

$$p(\mathbf{x}|\mathbf{w}_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mu_i) \right]$$

- Negative likelihood ratio.

$$\begin{aligned} & -\ln \frac{p(\mathbf{x}|\mathbf{w}_i)}{p(\mathbf{x}|\mathbf{w}_j)} \\ &= \frac{1}{2}(\mathbf{x} - \mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_i) - \frac{1}{2}(\mathbf{x} - \mu_j)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_j) \\ &= (\mu_j - \mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \frac{\mu_i + \mu_j}{2}) \end{aligned} \tag{22}$$

$$= [\mathbf{\Sigma}^{-1}(\mu_j - \mu_i)]^t \mathbf{x} + \frac{1}{2}(\mu_i - \mu_j)^t \mathbf{\Sigma}^{-1}(\mu_i + \mu_j) \tag{23}$$

$$= \mathbf{w}^t \mathbf{x} + w_0 \tag{24}$$

We obtain the FLD projection function $w = \mathbf{\Sigma}^{-1}(\mu_j - \mu_i)$, and the threshold is $w_0 = \frac{1}{2}(\mu_i - \mu_j)^t \mathbf{\Sigma}^{-1}(\mu_i + \mu_j)$. #