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:

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

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

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

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

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.

$$\begin{aligned}
& \max_{\mathbf{x}\in\mathbf{X},\ g(\mathbf{x})=0} f(\mathbf{x}) && (4)\\
= & \max_{\mathbf{x}\in\mathbf{X},\ g(\mathbf{x})=0} [f(\mathbf{x}) + \lambda^* g(\mathbf{x})] && (5)\\
\leq & \max_{\mathbf{x}\in\mathbf{X}} [f(\mathbf{x}) + \lambda^* g(\mathbf{x})] && (6)\\
= & [f(\mathbf{x}^*) + \lambda^* g(\mathbf{x}^*)] && (7)\\
= & f(\mathbf{x}^*) && (8)
\end{aligned}$$

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

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \lambda \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = 0 \tag{9}$$

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

$$g(\mathbf{x}) = 0 \tag{11}$$

Outline

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

Outline

2. Generalized Rayleigh Quotient

Fisher Criterion

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}} \tag{12}$$

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

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \tag{13}$$

for some constant λ.

2. Generalized Rayleigh Quotient

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

$$\max_{\mathbf{w}} \quad \mathbf{w}^t \mathbf{S}_B \mathbf{w} \tag{14}$$

$$s.t. \quad \mathbf{w}^t \mathbf{S}_W \mathbf{w} = K \tag{15}$$

which can be solved using Lagrange multipliers.

2. Generalized Rayleigh Quotient

Define the Lagrangian:

$$L = \mathbf{w}^t \mathbf{S}_B \mathbf{w} - \lambda(\mathbf{w}^t \mathbf{S}_W \mathbf{w} - K) \tag{16}$$

Maximize with respect to $\mathbf{w}$:

$$\nabla_{\mathbf{w}} L = 2(\mathbf{S}_B - \lambda \mathbf{S}_W)\mathbf{w} = 0 \tag{17}$$

To obtain the solution:

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w} \tag{18}$$

Outline

Outline

Chapter 4. Problem 35

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

Prove.

- Recall normal distribution.

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

- Negative likelihood ratio.

$$\begin{aligned}
& -\ln \frac{p(\mathbf{x}|\mathbf{w}_i)}{p(\mathbf{x}|\mathbf{w}_j)} \\
&= \frac{1}{2}(\mathbf{x}-\mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_i) - \frac{1}{2}(\mathbf{x}-\mu_j)^t \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_j) \\
&= (\mu_j - \mu_i)^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \frac{\mu_i+\mu_j}{2}) && (19) \\
&= \big[\mathbf{\Sigma}^{-1}(\mu_j-\mu_i)\big]^t \mathbf{x} + \frac{1}{2}(\mu_i-\mu_j)\mathbf{\Sigma}^{-1}(\mu_i+\mu_j)^{-1} && (20) \\
&= \mathbf{w}^t\mathbf{x} + w_0 && (21)
\end{aligned}$$

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