Computer Vision Hw2 – Problem3 & 4

1552746 Cui Hejie

May 14, 2018

1 Problem 3 (Math)

From the given conditions, we know that α_2 is an unit-vector, and among all the orientations orthogonal to α_1, the variance of data projections to α_2 is the largest one:

$$\alpha_2 = argmax(\alpha^T C\alpha) \ \ s.t. \ \alpha^T \alpha = 1, \alpha^T \alpha_1 = 0$$

From Lagrangian multiplier method, we can get that:

$$F = \alpha^T C\alpha - \lambda(\alpha^T \alpha - 1) - \mu\alpha^T \alpha_1$$

To maximum F, calculate the first derivative of F:

$$\frac{d(F)}{d(\alpha)} = 2C\alpha - 2\lambda\alpha - \mu\alpha_1 = 0(1)$$

Left multiply α_1^T on the equation:

$$2\alpha_1^T C\alpha - 2\lambda\alpha_1^T \alpha - \mu\alpha_1^T \alpha_1 = 2\alpha_1^T C\alpha - \mu = 0(2)$$

Since α_1 is a eigen vector of C, the corresponding eigen value of α_1 is λ_1, therefore,

$$C\alpha_1 = \lambda_1\alpha_1$$

So, equation (2) can be simplified as:

$$2\alpha_1^T C\alpha - \mu = 2(C\alpha_1)^T \alpha - \mu = 2(\lambda_1\alpha_1)^T \alpha - \mu = 2\lambda_1\alpha_1^T \alpha - \mu = 0$$

Since $\alpha_1^T \alpha = 0$, therefore,

$$\mu = 0$$

Together with equation (1), we know that,

$$C\alpha = \lambda\alpha$$

That is to say, the solution of the original equation is the eigenvector of the covariance matrix C Therefore(α is C's eigen-vector),

$$max(var(\alpha^T X)) = max(\alpha^T C\alpha) = max(\alpha^T \lambda\alpha) = max(\lambda)$$

We have known among all the orientations orthogonal to α_1, the variance of data projections to α_2 is the largest one, that is to say, α_2 corresponds to the largest eigen-value of vectors orthogonal to α_1. Besides, all eigen-vector is orthogonal to each other. So, α_2 actually is the eigen-vector of C associated to C's second largest eigen-value.

2 Problem 4 (Math)

Since each training sample is independent, first we prove that each single sample satisfy the following equation:

$$\nabla_\theta(-(y_i log(h_\theta(x_i)) + (1 - y_i)log(1 - h_\theta(x_i))) = x_i(h_\theta(x_i) - y_i)(1)$$

Function $\sigma(z) = \frac{1}{1+e^{-z}}$ is called sigmoid or logistic function, the probability that the testing sample x is positive is represented as:

$$h_\theta(x_i) = \sigma(\theta^T x_i)(2)$$

Find the gradient, the left side of equation (1) can be written as:

$$-(y_i \frac{\frac{dh_\theta(x_i)}{d\theta^T}}{h_\theta(x_i)} + (1-y_i)\frac{\frac{-dh_\theta(x_i)}{d\theta^T}}{1-h_\theta(x_i)})$$

Multiply x_i on the numerator and denominator:

$$-x_i(y_i \frac{\frac{dh_\theta(x_i)}{d\theta^T x_i}}{h_\theta(x_i)} + (1-y_i)\frac{\frac{-dh_\theta(x_i)}{d\theta^T x_i}}{1-h_\theta(x_i)})(3)$$

Take equation (2) into equation (3):

$$-x_i(y_i \frac{\frac{d\sigma(\theta^T x_i)}{d\theta^T x_i}}{\sigma(\theta^T x_i)} + (1-y_i)\frac{\frac{-d\sigma(\theta^T x_i)}{d\theta^T x_i}}{1-\sigma(\theta^T x_i)})(4)$$

One property of the sigmoid function:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}z} = \sigma(z)(1-\sigma(z))(5)$$

Take equation (5) into equation (4):

$$\begin{aligned}(4) &= -x_i(y_i \frac{\sigma(\theta^T x_i)(1-\sigma(\theta^T x_i))}{\sigma(\theta^T x_i)} + (1-y_i)\frac{-\sigma(\theta^T x_i)(1-\sigma(\theta^T x_i))}{1-\sigma(\theta^T x_i)}) \\ &= -x_i(y_i(1-h_\theta(x_i)) - (1-y_i)h_\theta(x_i)) \\ &= -x_i(y_i - h_\theta(x_i)) = x_i(h_\theta(x_i) - y_i)\end{aligned}$$

Sum up all the independent samples, we can find that for the cost function:

$$J(\theta) = -\sum_{i=1}^{m} y_i log(h_\theta(x_i)) + (1-y_i)log(1-h_\theta(x_i))$$

the gradient of this cost function is:

$$\nabla_\theta J(\theta) = \sum_{i=1}^{m} x_i(h_\theta(x_i) - y_i)$$