CS28010 Homework 1

Guoxin SUI

October 27, 2017

1 Minimizing error

1.1 Point representation

Suppose we have N data points $x_i, i = 1, ..., N$. Please find one single point to best represent these N data points.

For a point x, we define the cost function as $J(x) = \frac{1}{2}\sum_{i=1}^{N}(x - x_i)^2$. For the best point x^*, we have $\frac{dJ(x^*)}{dx^*} = 0$

Then we have:

$$\begin{aligned}
\frac{dJ(x^*)}{dx^*} &= \frac{\sum_{i=1}^{N}(x^{*2} - 2x^* x_i + x_i^2)}{2dx} \\
&= \sum_{i=1}^{N}(x^* - x_i) \\
&= mx^* - \sum_{i=1}^{N}(x_i) \\
&= 0
\end{aligned}$$

So

$$x^* = \frac{1}{m}\sum_{i=1}^{N}(x_i)$$

which is the mean of these N points.

1.2 Line representation

Suppose we have N pairs of data tuples: $(x_i, y_i), i = 1, ..., N$, where x_i is a two dimensional vector $[x_{i1}, x_{i2}]^T$. Now we want to fit a line of form $y = w^T x + b + e$ to represent these N data tuples, where e is error. Please find the best **w** and b. You can use the methods you learned in high school to solve this problem. And bonus points will be given to students who solve this problem by matrix calculus.

We define

$$
\begin{aligned}
X &= \left[(1,x_1)^T, (1,x_2)^T, ...(1,x_N)^T\right], \\
\vec{y} &= \left[y_1, y_2, ...y_N\right], \\
\theta &= \left[b, w\right], \\
J(x) &= \frac{1}{2}\sum_{i=1}^{N}(X_i - \vec{y_i})^2
\end{aligned}
$$

To minimize J, we take its derivatives with respect to θ. Hence,

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta \frac{1}{2}\sum_{i=1}^{N}(X_i - \vec{y_i})^2 \\
&= \nabla_\theta \frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y}) \\
&= \frac{1}{2}\nabla_\theta(\theta^T X^T X\theta - \theta^T X^T \vec{y} - \vec{y}^T X\theta + \vec{y}^T\vec{y}) \\
&= \frac{1}{2}\nabla_\theta tr(\theta^T X^T X\theta - \theta^T X^T \vec{y} - \vec{y}^T X\theta + \vec{y}^T\vec{y}) \\
&= \frac{1}{2}\nabla_\theta(tr\theta^T X^T X\theta - 2tr\vec{y}X\theta) \\
&= \frac{1}{2}(X^T X\theta + X^T X\theta - 2X^T\vec{y}) \\
&= X^T X\theta - X^T\vec{y}
\end{aligned}
$$

We set the derivatives to zero, then we have

$$\theta = (X^T X)^{-1} X^T \vec{y}$$

For this question,

$$
\begin{aligned}
w &= \theta[1:] \\
b &= \theta[0]
\end{aligned}
$$

2 Separating Boundary

Suppose we have two Gaussian distributions for two different classes of data $N(x|\mu_1, \Sigma_1)$ and $N(x|\mu_2, \Sigma_2)$, where x is a two-dimensional vector. For all x_0 that satisfy $N(x_0|\mu_1, \Sigma_1) = N(x_0|\mu_2, \Sigma_2)$, we call these x_0 as lying on the separating boundary. (We assume these two classes have the same priors)

2.1 Line boundary

Suppose $N(x|\mu_1, \Sigma_1) = \frac{1}{2\pi|\Sigma_1|^{1/2}} exp(-\frac{1}{2}(x-\mu_1)^T {\Sigma_1}^{-1}(x-\mu_1)$, where

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\Sigma_1 = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$$

Please find all settings of μ_2 and Σ_2 that makes a straight line boundary between the two classes.

Initiate LH Algorithm:

2.2 Other forms of boundary

Discuss the conditions where the separating boundaries between the two classes are parabola, hyperbola, ecllipse and line. Tips: you may want to refer to `https://en.wikipedia.org/wiki/Conic_section`.

Initiate LH Algorithm:

2.3 Optional summary work

Note: this is an optional homework. Please give your understanding of the reason why error terms often subject to a Gaussian distribution. Students who complete this part will get bonus points.

In probability theory, the central limit theorem (CLT) establishes that, in most situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

In my understanding, when the error terms are independent, they can be considered as a small variable who is randomly distributed, it is reasonable to subject to a Gaussian distribution.