Tutorial for A2

Thursday, September 18, 2014 3:02 PM

mean of the first class

$$\left[\mu_1, \mu_2, \sigma \; \theta\right] = u$$

$$u_1 = \frac{1}{N}\sum_{i=1}^{N} x_i$$

only for points that belong to class "such that $y_i = 1$"

$$\log \; P(D \mid \mu_1, \mu_2 \, \sigma^2)$$

$$f(u) \; \sum \; I(y_i = 1) \; P(\quad)$$

$$I(y_i = 2 \; P(\quad)$$

function of the x_i's

$$\frac{df}{d\mu_1} = 0$$

2.2| given

random variable

$$x \sim N(\mu, \sigma^2)$$

$$\mu \sim N(\alpha, \gamma^2)$$

$$p(\mu \mid \underbrace{x_i, \alpha, \sigma^2, \gamma^2}_{\theta})$$

bayes rule

$$\propto \; p(x \mid \mu, \theta) \; p(\mu \mid \theta)$$

gaussian, so find mean and variance

$$p(\mu \mid x, \theta) \propto \lambda x \;\; p\left(-\frac{\lambda(x-\mu)^2}{2}\right)$$

$$* \; \exp\left(-\lambda \; \frac{(\mu-\alpha)^2}{2}\right)$$

don't need to find denominator, so we can ignore the constants (constant with respect to μ)

$$\exp\left(-\frac{1}{2b^2}(\mu - a)^2\right)$$

$$\leftarrow \; \sim N(a, b^2) \;\Rightarrow\; \frac{1}{\sqrt{2\pi}\, b} \exp\left(-\frac{(\mu-a)^2}{2b^2}\right)$$

2.3| different diagonal elements

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

d by 1

$$f(x) = x^T a$$

d by 1

$$\nabla f(x) = \overset{d \text{ by } 1}{a}$$

$$f(x) = a^T x \;\Rightarrow\; x^T a$$

$$\nabla f(x) = a$$

1 by d, d by 1

$$f(x) = x^T A x \qquad x^T A^T x$$

$$\nabla f(x) = Ax + A^T x$$

$\leftarrow$ in general A is not senetric

3.3

$$g: \mathbb{R}^n \to \mathbb{R}$$

find gradiant, and Hessian

$$f(\underset{d \text{ by } 1}{x}) = g(\overset{n \text{ by } d}{A}x)$$

$x: \mathbb{R}^d$

$$\underset{d \text{ by } 1}{\nabla f(x)} \qquad \underset{n \text{ by } 1}{\nabla g}(\underbrace{Ax}_{n \text{ by } 1})$$

$$\underset{d \text{ by } 1}{\nabla f(x)} = \underset{d \text{ by } n}{A^T} \; \underset{n \text{ by } 1}{\nabla g(Ax)}$$

Scalar:

$$h(x) = g(ax)$$

$$\frac{d}{dx} = g'(ax)\,a$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial x}{\partial x_1} \\ \vdots \\ \frac{\partial x}{\partial x_d} \end{bmatrix}$$

(vector ~~of~~ partials $x_1 \cdots x_d$)

$$\frac{\partial^2 f}{\partial x_i x_j}$$

$$\begin{bmatrix} 11 & 12 & 1d \\ 21 & & \\ d1 & & dd \end{bmatrix}$$

$$\overset{d \text{ by } d}{H} = \nabla\left[(\nabla f(x))^T\right]$$

$$= \nabla\left(\nabla^T g(Ax)\,A\right)$$

$$= \underset{d \text{ by } n}{A^T} \underbrace{\nabla^2 g(Ax)}_{n \text{ by } n} \underset{n \text{ by } d}{A}$$

4.2

$$\underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{z_i}{2}(w^T x_i - y_i)^2$$

convert to matrix form

$$= (Xw - Y)^T \underset{\text{diagonalmatrix of } z_i}{Z} (Xw - Y)$$

$$\begin{bmatrix} z_1 & & & 0 \\ & z_2 & & \\ & & \ddots & \\ 0 & & & z_i \end{bmatrix}$$

4.3

$$\underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} |w^T x_i - y_i|$$

(not differentiable)

rewrite as an objective function with linear constraints, and solve as a linear program

$$\underset{w,V}{\operatorname{argmin}} \sum_{i=1}^{N} V_1 \qquad |w^T x_i - y_i| \le V_i$$

minimize upperbound

$$\Downarrow$$

$$-V_i \le w^T x_i - y_i \le V_i$$

$$\Downarrow$$

constraints: $\begin{cases} w^T x_i - y_i \le V_i \\ w^T x_i - y_i \ge -V_i \end{cases}$

$$\Downarrow$$

$$\begin{cases} w^T x_i - V_i \le y_i \\ -w^T x_i - V_i \le -y_i \end{cases}$$

$\nwarrow$ we have $\overset{N}{i=1}$ of these

$$\underset{\theta}{\min} \; f\,\theta$$

s.t. $A\theta \le b$

$A_{eq}\,\theta = b_{eq}$

$$\underset{\text{lower}}{l} \le \theta \le \underset{\text{upper}}{p}$$

$$\theta = [\, w_{1:d} \quad V_{1:n} \,]$$

$$f = [\, \underset{d \text{ zeros}}{\text{zeros}} \quad \underset{n \text{ ones}}{\text{ones}} \,]$$

matrix form $\begin{cases} Xw - V \le Y \\ -Xw - V \le -Y \end{cases}$

$$A = \left[\begin{array}{c|c} \overset{Xw}{X} & \overset{\text{to get } -V}{-I} \\ \hline \underset{-Xw}{-X} & \underset{-V}{-I} \end{array}\right]$$

$$b = \left[\begin{array}{c} Y \\ \hline -Y \end{array}\right]$$

lower bound

$$l = [\, \underset{1\times d}{-\infty} \; \cdots \; 0 \cdots]$$

$$[\, d\;(-\infty)'s \; \cdots \; N \text{ zeroes} \,]$$