4/25/18

Note Title 1/20/2000

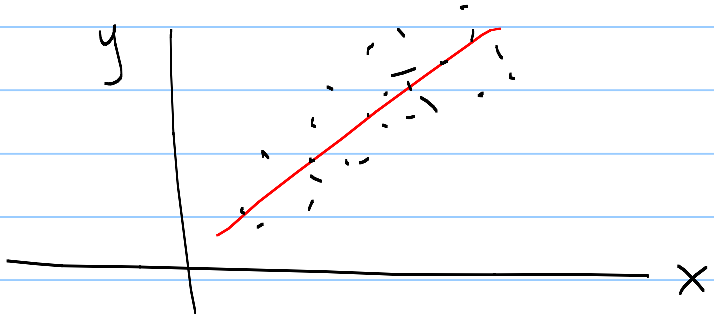
Stochastic Approaches to Linear Regression

Regression data

$$\{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \ldots, (\bar{x}_p, y_p)\}, \quad x \in R^N$$

We model directly the distribution $P_r(y/\bar{x})$
(discriminative methods)

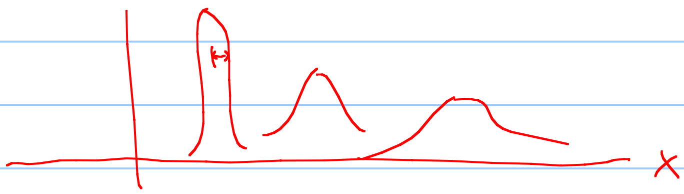
Goal: predict the posterior distribution



$$\therefore \boxed{P_r(y_p \mid \bar{x}_p, \bar{\theta}) = \text{Norm}_{y_p}\left[b + \bar{x}_p^T \bar{w}, \sigma^2\right], \quad \bar{\theta} = \{b, \bar{w}, \sigma^2\}}$$

Univariate normal or Gaussian

$$P_r(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = \text{Norm}_x\left[\mu, \sigma^2\right]$$



$$\tilde{x}_p = [1 \ \bar{x}_p]^T, \quad \tilde{w} = [b \ \bar{w}]^T$$

$$P_r(y_p \mid \tilde{x}_p, \bar{\theta}) = \text{Norm}_{y_p}\left[\tilde{x}_p^T \tilde{w}, \sigma^2\right]$$

Learning

Find $\bar{\theta} = \{\tilde{w}, \sigma^2\}$

Multivariate Normal

$(x \in \mathbb{R}^N)$ $$Pr(\bar{x}) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma^{-1} (\bar{x}-\mu)\right]$$

$$= Norm_{\bar{x}}[\bar{\mu}, \Sigma]$$

- covariance matrix Σ, $N \times N$, symmetric, positive definite

examples (2D)

$$\Sigma_{spher} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I, \quad \Sigma_{diag} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}, \quad \Sigma_{full} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$$

$\sigma_{12} = \sigma_{21}$

- When the covariance is spherical or diagonal, the individual variables are independent

e.g. 2D diagonal

joint $\rightarrow$ $$Pr(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x_1 \; x_2) \Sigma^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right]$$

$|\Sigma| = \sigma_1^2 \cdot \sigma_2^2$

$$= \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x_1 \; x_2) \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[\frac{-x_1^2}{2\sigma_1^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[\frac{-x_2^2}{2\sigma_2^2}\right]$$

$$= Pr(x_1) \cdot Pr(x_2)$$

Form matrices & vector

$$\bar{y} = [y_1, .., y_P]^T, \quad \tilde{X} = [\tilde{x}_1 \; \tilde{x}_2 .. \tilde{x}_P] = \begin{bmatrix} 1 & 1 & & 1 \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_P \end{bmatrix}_{(N+1)\times P}$$

$$\tilde{w} = [b \; w_1 \cdots w_N]^T_{(N+1)\times 1}$$

$$Pr(\bar{y} \mid \tilde{X}, \bar{\theta}) = Norm_{\bar{y}}[\tilde{X}^T\tilde{w}, \sigma^2 I]$$

Learning

$$\hat{\bar{\theta}} = \arg\max_{\bar{\theta}} [Pr(\bar{y} \mid \tilde{X}, \bar{\theta})] = \arg\max_{\bar{\theta}} \{\log[Pr(\bar{y} \mid \tilde{X}, \bar{\theta})]\}$$

$$\hat{\tilde{w}}, \hat{\sigma}^2 = \underset{\tilde{w}, \sigma^2}{\arg\max} \left\{ \log\left[\frac{1}{(2\pi)^{P/2}(\sigma^2)^{P/2}} \exp\left[-\frac{1}{2\sigma^2}(\bar{y} - \tilde{X}\tilde{w})^T(y - \tilde{X}\tilde{w})\right]\right]\right\}$$

$$= \underset{\tilde{w}, \sigma^2}{\arg\max} \left\{ \underbrace{-\frac{P}{2}\log(2\pi) - \frac{P}{2}\log(\sigma^2) - \frac{(\bar{y} - \tilde{X}\tilde{w})^T(y - \tilde{X}\tilde{w})}{2\sigma^2}}_{a} \right\}$$

$$\nabla_{\tilde{w}} a = 0$$

$$\Rightarrow 2\tilde{X}\tilde{X}^T\tilde{w} - 2\tilde{X}\bar{y} = 0 \Rightarrow \tilde{X}\tilde{X}^T\tilde{w} = \tilde{X}\bar{y}$$

$$\Rightarrow \boxed{\tilde{w} = (\tilde{X}\tilde{X}^T)^{-1}\tilde{X}\bar{y}}$$

if $\tilde{X}\tilde{X}^T$ invertible

EXACTLY the same expression we derived using least-squares.

$$\frac{da}{d\sigma^2} = 0$$

$$\Rightarrow -\frac{P}{2}\frac{1}{\sigma^2} + \frac{(\bar{y} - \tilde{X}\tilde{w})^2}{2\sigma^4} = 0 \Rightarrow \boxed{\sigma^2 = \frac{(\tilde{y} - \tilde{X}\tilde{w})^2}{P}}$$

3 issues

1. predictions over-confident
2. only linear function
3) all $\bar{x}$ are involved