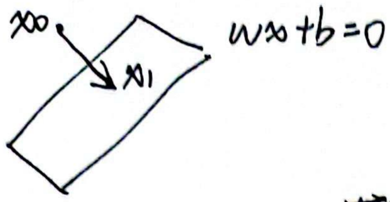
Chapter 6. P141

6-1



法向量 (normal vector) w.

$\overrightarrow{x_0x_1}$ intersect $wx+b=0$ with x_1.

~~列方程~~ $\begin{cases} x_1 = x_0 + kw, \\ w^Tx_1 + b = 0 \end{cases}$

$$\Rightarrow w^Tx_0 + kw^Tw + b = 0$$

$$\Rightarrow k = \frac{-(w^Tx_0+b)}{\|w\|_2}, \quad len = \frac{w^Tx_0+b}{\|w\|_2}.$$

6.4. 线性判别分析 (LDA):

$$\begin{cases} \min\limits_{w} \; -w^TS_b \cdot w \\ s.t. \; w^TS_ww = 1 \end{cases}, \qquad S_b = (\mu_0-\mu_1)(\mu_0-\mu_1)^T$$

$$S_w = \Sigma_0 + \Sigma_1$$

linear SVM:

assume ~~$\beta = (w; b)$~~, $b=0$.

$$\min_{w,b} \|w\|_2 \quad s.t. \; y_i(w^Tx_i+b) \geq 1 \quad \Rightarrow \quad \cancel{\min_{\beta}} \; \min \|w\|_2 \quad s.t. \; y_iw^Tx_i \geq 1$$

$$w^TS_ww = Var(w^TX_0) + Var(w^TX_1) = Var(w^T(X_0+X_1)) = 1$$

~~$= Var(w^TX) = 1$~~

假设 X_0, X_1 一样多

$$\min_{w} -w^TS_bw = \min_{w} -(E(w^TX_0) - w^TX_1))^2$$

, ~~assume X_0, X_1~~ ~~set~~ ~~是~~ ~~X_0+X_1~~ ~~have same terms~~

equivalent to $\min\limits_{w} E(w^TX)^2$

$s.t. \; Var(w^TX) = 1$

LDA 与所有点有关，

SVM 与支持向量有关，故当所有点为支~~撑~~持向量时，

LDA = SVM.

6.6 SVM 只对 support vector 有关

support vector ~~指~~ 为 ~~在对~~ $y_i(w^T x_i + b) \le 1$ 的向量，

即噪声影响很大.

6.9. logistic regression: P59

minimize $\ell(\beta) = \sum_{i=1}^{m} -\ln\left(y_i p_1(\hat{x}_i;\beta) + (1-y_i) p_0(\hat{x}_i;\beta)\right)$

$$= \sum_{i=1}^{m} \left(-y_i \beta^T \hat{x}_i + \ln(1+e^{\beta^T \hat{x}_i})\right)$$

?

$$\min_{\beta} \ell(\beta) \sum_{i=1}^{m} -y_i \beta^T (\hat{x}_i) + \ln(1+e^{\beta^T (\hat{x}_i)})$$

set $h(x_i) = \beta^T \hat{x}_i$

$$= \min_{h(x_i)} \sum_{i=1}^{m} -y_i h(x_i) + \ln(1+e^{h(x_i)})$$

根据 representer thm, $\exists\ h^*(x) = \sum_{i=1}^{m} \alpha_i K(x, x_i)$ 为解形式

$$\min_{h(x_i)} \sim = \min_{\alpha} \sum_{i=1}^{m} \left[-y_i \sum_{j=1}^{m} \alpha_j K(x_i, x_j) + \ln\left(1+\prod_{j=1}^{m} e^{\alpha_j K(x_i, x_j)}\right)\right]$$

则 α 对应原问题 β，$K(x_i, x_j)$ 对应 $\hat{x}_i$，

~~得到~~ 可通过 Newton descent method 得到近似解. □

6.10，可以先用 kernel LDA 方法，把偏差大的数据排除，

再用剩下数据做 kernel SVM. ✓