Assignment 2
Solutions

CS - 6375 . 501

Machine Learning

Assignment 2 Solutions

1
a) $$\min_{w} \sum_{i=1}^{M} (y_i - w^T x_i)^2 + \lambda \|w\|^2$$

$$\Rightarrow \min_{w, \xi_i} \sum_{i=1}^{m} \xi_i^2 + \lambda \|w\|^2$$

$$\text{s.t.} \quad y_i - w^T x_i = \xi_i$$

$$L(w, \xi, \alpha) = \sum_{i=1}^{m} \xi_i^2 + \lambda \|w\|^2 + \sum_{i=1}^{m} \alpha_i (y_i - w^T x_i - \xi_i) \quad \text{--- } (1)$$

$$\frac{\partial L}{\partial \xi_i} = 2\xi_i - \alpha_i = 0$$

$$\Rightarrow \xi_i = \alpha_i / 2 \quad \text{--- } (2)$$

$$\frac{\partial L}{\partial w} = 2\lambda w - \sum_{i=1}^{m} \alpha_i x_i = 0$$

$$\Rightarrow w = \frac{\sum_{i=1}^{m} \alpha_i x_i}{2\lambda} \quad \text{--- } (3)$$

Substitute (2) & (3) into (1)

Note: No inequality constraint: $\alpha_i \geq 0$

Final Expression

$$G(\alpha) = \sum_i -\frac{\alpha_i^2}{4} - \frac{1}{4\lambda} \sum_{i,j} \alpha_i \alpha_j x_i^T x_j$$

$$+ \sum_i \alpha_i y_i$$

b) $$\min_{w,b} \sum_{i=1}^{M} \max(0, 1 - y_i(w^T x_i + b))^2 + \lambda \|w\|^2$$

$$\Rightarrow \min_{w, b, \varepsilon_i} \sum_{i=1}^{M} \varepsilon_i^2 + \lambda \|w\|^2$$

$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1 - \varepsilon_i$$

$$\varepsilon_i \geq 0$$

$$L(\alpha, \beta, w, b, \varepsilon_i) = \sum_{i=1}^{M} \varepsilon_i^2 + \lambda \|w\|^2$$

$$+ \sum_{i=1}^{M} \alpha_i (1 - \varepsilon_i - y_i(w^T x_i + b)) - \beta_i \varepsilon_i$$

1

$$\frac{\partial L}{\partial \varepsilon_i} = 2\varepsilon_i - \alpha_i - \beta_i$$

$$\Rightarrow \varepsilon_i = (\alpha_i + \beta_i)/2 \quad — \quad 1$$

$$\nabla_w L = 2\lambda w - \sum_{i=1}^{N} y_i x_i = 0$$

$$\Rightarrow w = \frac{\sum_{i=1}^{M} y_i x_i \alpha_i}{2\lambda} \quad — \quad 3$$

$$\nabla_b L = \sum_i \alpha_i y_i = 0 \quad — \quad 4$$

substitute 2 & 3 into 1 with constraint 4 & $\alpha_i \geq 0, \beta_i \geq 0$

<u>Final Expression</u>

$$\max \left(- \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \xrightarrow{} (\alpha_i + \beta_i)^2 + 4\lambda \sum_{i=1}^{M} \alpha_i \right.$$

$$\text{s.t} \quad \sum_{i=1}^{M} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0$$

(c) First compute b using complementary slackness.

$$b = y_i - w^T \phi(x_i) \quad , \text{ for } i : \lambda_i > 0$$

$$= y_i - \sum_j \lambda_j y_j \phi(x_j)^T \phi(x_i)$$

$$= y_i - \sum_j \lambda_j y_j k(x_i, x_j)$$

Next, $y_t = w^T \phi(x_t) + b$

$$= \sum_j y_j \lambda_j k(x_i, x_t) + b$$

* No need to compute w explicitly

Substitute $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

to get expression

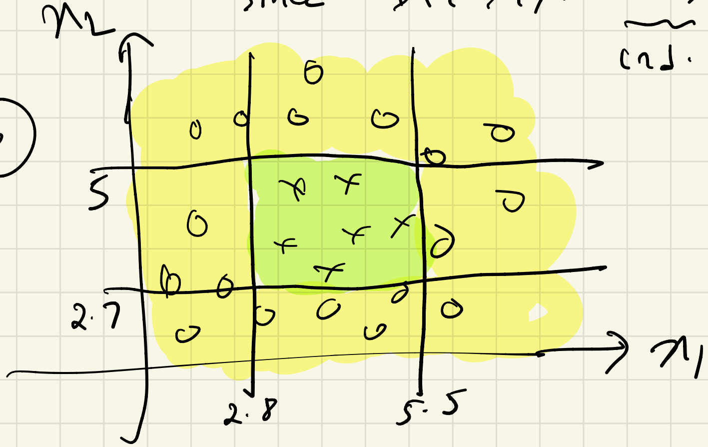
(2)

(a) - Give an example (see example in class slides) which shows that maximizing $I(X_i ; Y)$ selects features which split the dataset based on label

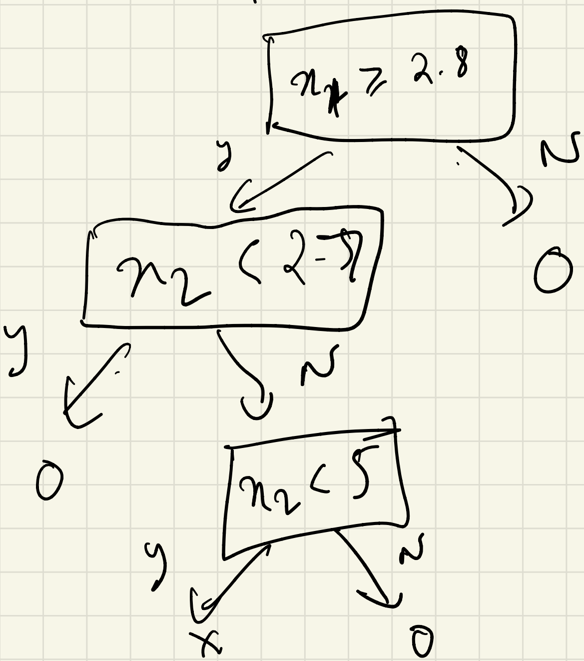
- maximizing $I(X_i ; Y) \equiv$ minimizing $H(Y|X_i)$

since $I(X_i ; Y) = \underbrace{H(Y)}_{\text{ind. of } X_i} - H(Y|X_i)$

(b)



One possible Decision Tree.



(c) A & C are true.

(3) $$P(\mu \mid x^1, \ldots, x^N) = \frac{P(x^1, \ldots x^N \mid \mu)\, P(\mu)}{P(x^1, \ldots, x^N)}$$

Now, $$P(x^1, \ldots, x^N \mid \mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^i-\mu)^2}{2\sigma^2}}$$

$$P(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(\mu-\nu)^2}{2\beta^2}}$$

$$\Rightarrow \log P(\mu \mid x^1, \ldots, x^N) = \sum_{i=1}^{N}\left[-\log\sqrt{2\pi\sigma^2} - \frac{(x_i-\mu)^2}{2\sigma^2}\right] - \log\sqrt{2\pi\beta^2} - \frac{(\mu-\nu)^2}{2\beta^2}$$

$$\Rightarrow \frac{\partial \log P(\mu \mid x^1, \ldots, x^N)}{\partial \mu} = \sum_{i=1}^{N} \frac{(x_i-\mu)}{\sigma^2} - \frac{(\mu-\nu)}{\beta^2} = 0$$

$$\Rightarrow \frac{\mu-\nu}{\beta^2} = \sum_{i=1}^{N} \frac{x_i}{\sigma^2} - \frac{N\mu}{\sigma^2}$$

$$\Rightarrow \frac{\mu}{\beta^2} + \frac{N\mu}{\sigma^2} = \frac{\sum_{i=1}^{N} x_i}{\sigma^2} + \frac{\nu}{\beta^2}$$

$$\Rightarrow \mu_{MAP} = \frac{\sigma^2\nu + \beta^2 \sum_{i=1}^{N} x_i}{\sigma^2 + N\beta^2}$$

As $N \to \infty$, $\mu_{MAP} \to \mu_{MLE} = \dfrac{\sum_{i=1}^{N} x_i}{N}$