Final Review (contied)

3 Classification

3.1 Perceptron : an algorithm for learning a binary classifier of the form

$$h(x) = \text{sign}(w^T x + b)$$

where $w \in \mathbb{R}^p$ is the weight and $b \in \mathbb{R}$ is the bias. (WLOG, $b=0$)

Setting: Given labeled sample pts $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ with $y_i = \pm 1$.

If $\exists i$ s.t. $y_i \left(x^{(i)T} w^{(k)} \right) < 0$,

update $w^{(k+1)} = w^{(k)} + y_i x^{(i)}$.

Algorithm: Lecture 26

Related Topics: convergence under linear separability assumption. (HW8 #1)

3.2 **SVM**: a linear binary classifier based on margin maximization.

Setting: Given labeled sample pts $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ with $y_i = \pm 1$, solve the convex optimization problem:

$$\min_{w \in \mathbb{R}^p} \|w\|^2$$

$$\text{s.t. } y_i \left(x^{(i)T} w + b \right) \geq 1, \quad i = 1, \cdots, n.$$

$$\left(\text{margin} = \frac{2}{\|w\|^2} \right)$$

3.3 LDA: a dimension reduction method that maximizes projected data separability.

Setting: Given labeled sample pts
$(x^{(1)}, y_1), \dots, (x^{(n)}, y_n)$ with $y_i = \pm 1$,

two classes: $\mathscr{C}_1 = \{x^{(i)} : y_i = 1\}$ $\quad$ $\mathscr{C}_2 = \{x^{(i)} : y_i = -1\}$

# of sample pts: n_1 $\quad$ n_2

means $\mu^{(1)} = \frac{1}{n_1} \sum_{x^{(i)} \in \mathscr{C}_1} x^{(i)}$ $\quad$ $\mu^{(2)} = \frac{1}{n_2} \sum_{x^{(i)} \in \mathscr{C}_2} x^{(i)}$

sample covariance matrix

$$C^{(1)} = \frac{1}{n_1} \sum_{x^{(i)} \in \mathscr{C}_1} (x^{(i)} - \mu^{(1)})(x^{(i)} - \mu^{(1)})^T \quad C^{(2)} = \frac{1}{n_2} \sum_{x^{(i)} \in \mathscr{C}_2} (x^{(i)} - \mu^{(2)})(x^{(i)} - \mu^{(2)})^T$$

Solve $\arg\max_{\|u\|=1} \frac{u^T S_B u}{u^T S_W u}$ $\left(\begin{array}{l} = \text{eigenvector} \\ \text{associated to} \\ \text{the largest eigenvalue} \\ \text{of } S_W^{-1} S_B \end{array}\right)$

where $S_B \stackrel{\text{def}}{=} (\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})^T$ is the between-class scatter matrix.

and $S_W \stackrel{\text{def}}{=} n_1 C^{(1)} + n_2 C^{(2)}$ is the within-class scatter matter.

Algorithm : Lecture 32

Related Topics : generalized Rayleigh quotient (HW9 #2 #3)

4 Linear Methods in Regression: Statistical Point of View.

- Formulation of three linear models with noise
- Def of bias, variance, MSE and their relations (bias-variance tradeoff)

$$MSE(\beta) = \|Bias(\beta)\|^2 + tr\, Cov(\beta)$$

- Def and derivation of β^{blue}, β^{map}, and their relations with β^{ls}, β^{ridge}, β^{lasso}.

  (HW10 #3 #4)