머신러닝 강의필기 11주차 김선수 컴퓨터공학과 2016104109.

* 널리 쓰이는 커널 함수.

- Polynomial : $k(x,z) = (x \cdot z + 1)^p$
- Gaussian : $k(x,z) = \exp\left(\frac{-\|x-z\|_2^2}{2\sigma^2}\right)$
- Hyperbolic Tangent : $k(x,z) = \tanh(\alpha x \cdot z + \beta)$

* SVM 최적화 알고리즘

SMO (Sequential Minimal Operation)

→ $f = 2$, 두 개의 Lagrange Multiplier 구하기

Cutting-Plane

→ SMO보다 faster.

* SVM Prediction (linear)

Weight 계산 : $w = \sum_{i=1}^{n} \alpha_i y_i x_i = \sum_{\alpha_k \neq 0} \alpha_k y_k x_k$

Bias 계산 : $b = y_i - w^T x_i$

Output 계산 : $d(x) = w^T x + b \gtrless 0$ $\begin{cases} out = 1 \\ out = -1 \end{cases}$

* SVM Prediction (non-linear).

Weight 계산 : $w = \sum_{i=1}^{n} \alpha_i y_i x_i$

Bias 계산 : $b_k = y_i - \sum_{k=1}^{n} \alpha_k y_k k(x_k, x_i)$

Output 계산 : $d_k(x) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b \gtrless 0$.

$\begin{cases} out = 1 \\ out = -1 \end{cases}$

* Multi-Class SVM.

- 이진 분류기 C개를 class k과 나머지 C-1개 c
- $d_k(x)$가 가장 큰 값을 갖는 class.

$\hat{k} = \arg\max_k d_k(x)$.

* SVM Regression

+ Training : $X = \{(x_1, y_1) \cdots (x_n, y_n)\}$

+ Linear Regression Function.

→ $f(x) = w_1 x_1 + \cdots + w_d x_d + b = x^T w + b$.

+ Regularized Cost Function

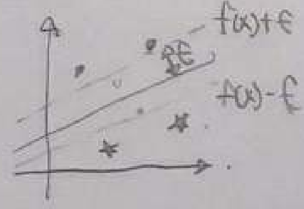
→ $J(w) = C \sum_{i=1}^{n} \| y_i - f(x_i) \| + \frac{1}{2} \|w\|_2^2$

* SVM Constrained Minimization

→ $J(w, \xi) = \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{n} \xi_i$

→ $y(w^T x_i + b) \geq 1 - \xi_i$ → $|y_i - (w^T x_i + b)| \leq$

+ ε-둔감 오류함수 → 비선형 확장.



○ → $f(x_i) - \epsilon \leq y_i \leq f(x_i) +$

● → $y_i - f(x_i) > \epsilon$

☆ → $y_i - f(x_i) < -\epsilon$

⇒ Lagrangian, KKT 조건, 커널 trick 적용

$L(\alpha, \tilde{\alpha}) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \tilde{\alpha}_i)(\alpha_j - \tilde{\alpha}_j) k(x_i, x_j)$

minimize. $- \epsilon \sum_{i=1}^{n} (\alpha_i + \tilde{\alpha}_i) + \sum_{i=1}^{n} (\alpha_i - \tilde{\alpha}_i) y_i$

Subject to $\sum_{i=1}^{n} (\alpha_i - \tilde{\alpha}_i) = 0$.