Note 1

Lai Zehua 2014012668

2017 10 14

VC-Dimension. *If $\exists x_1, ..., x_n, s.t. \mathcal{F}$ shatters $x_1, ..., x_n$ and $\forall x_1, ..., x_{n+1}$, $\mathcal{F}$ can not shatter $x_1, ..., x_{n+1}$. n is the **VC-dimension** of $\mathcal{F}$*

Linear Classifier.

$$max_{w,b,t} t$$
$$s.t. y_i(w^T x_i + b) \geq t$$
$$\|w\| = 1$$

or

$$min_{w,b} \frac{1}{2}\|w\|^2$$
$$s.t. y_i(w^T x_i + b) \geq 1$$

Appendix: Game Theory & Lagrange Duality. *In two player zero-sum game, We have minmax = maxmin in a Nash equilibrium.*[1]

$$\max_{y \in A_2} \min_{x \in A_1} u_1(x, y) = \min_{x \in A_1} \max_{y \in A_2} u_1(x, y)$$

Problem 1. *For a matrix M, $\min_i \max_j M_{ij} \geq \max_j \min_i M_{ij}$ which means that if two player choose pure strategy and to act sequentially, second one gets an advantage.*

If two players can choose mixed strategy, a Nash equilibrium exists, so the equation above holds.

Saddle Point Theorem. *$f(x, y)$ Fix y, $f(\cdot, y)$ is convex. Fix x, $f(x, \cdot)$ is concave, then*

$$\max_y \min_x f(x, y) = \min_x \max_y f(x, y)$$

If $f(x), g_i(x)$ are convex and $h_i(x)$ is linear, the optimization problem

$$\begin{cases} minf(x) \\ s.t. g_i(x) \leq 0 \\ h_i(x) = 0 \end{cases}$$

$$\iff \begin{cases} min_x max_{\mu,\lambda} f(x) + \sum \mu_i h_i(x) + \sum \lambda_i g_i(x) \\ s.t. \lambda_i \geq 0 \end{cases}$$

$$\iff \begin{cases} max_{\mu,\lambda} min_x f(x) + \sum \mu_i h_i(x) + \sum \lambda_i g_i(x) \\ s.t. \lambda_i \geq 0 \end{cases}$$

Let $L(x, \mu, \lambda) = f(x) + \sum \mu_i h_i(x) + \sum \lambda_i g_i(x)$, solve $\frac{\partial L}{\partial x} = 0$ to get $x^* = \phi(\mu, \lambda)$. It suffices to solve $max_{\mu,\lambda} f(\phi(\mu, \lambda)) + \sum \mu_i h_i(\phi(\mu, \lambda)) + \sum \lambda_i g_i(\phi(\mu, \lambda))$.

The linear classifier problem is $\begin{cases} min_{w,b} \frac{1}{2}\|w\|^2 \\ s.t. y_i(w^T x_i + b) \geq 1 \end{cases}$

[1] A course in Game Theory, P25

$$\iff \begin{cases} min_{w,b,\lambda_i} L(w,b) = \frac{1}{2}\|w\|^2 - \sum \lambda_i [y_i(w^T x_i + b) - 1] \\ \lambda_i \geq 0 \end{cases}$$

$L(w,b) = \frac{1}{2}\|w\|^2 - \sum \lambda_i [y_i(w^T x_i + b) - 1]$, $w^* = \sum \lambda_i y_i x_i$, $\sum \lambda_i y_i = 0$
KKT-conditions. *(1) Stationary:* $\nabla L(x,\lambda,\mu)|_{x^*,\lambda^*,\mu^*} = 0$ *(2) Primal feasibility:* $h_i(x^*) = 0, g_i(x^*) \leq 0$ *(3) Dual Feasible* $\lambda^* \geq 0$ *(4)Complementary slackness* $\lambda_i g_i(x^*) = 0$

Problem 2. *KKT is necessary condition.*

Problem 3. *If $f(x), g_i(x)$ are convex and $h_i(x)$ is linear, KKT is sufficient condition.*

By KKT condition, $\lambda_i^*[y_i(w^{*T} x_i + b^*) - 1] = 0$. $\lambda_i^* = 0$ for all (x_i, y_i) that are not closest to the hyperspace. $\lambda_i^* \neq 0$ for all support vector. It is the Support Vector Machine (SVM).
Term Project. *Data space = [N],* $\mathcal{F} \subseteq \{0,1\}^N$*, VC dmension* $VC(\mathcal{F}) = d$ *if* $\exists i_1, ..., i_d$*,* $\mathcal{F}$*'s projection onto* $i_1, ..., i_d$ *contains* $\{0,1\}^d$ *and* $\forall i_1, ..., i_{d+1}$*'s projection onto* $i_1, ..., i_{d+1} \neq \{0,1\}^{d+1}$

$f \in \mathcal{F}$*,* $\exists X_f \subseteq X$*,* $f|_{X_f} \neq f'|_{X_f}, \forall f' \in \mathcal{F}$

Teaching dimension of f is $TD(f,\mathcal{F}) = min|X_f|$*, best case teaching dimension is* $TD_{min}(\mathcal{F}) = min_{f\in\mathcal{F}} TD(f,\mathcal{F})$.

Let $\mathcal{F}_0 = \mathcal{F}$ *and assume* $f_1, s.t. TD_{min}(\mathcal{F}) = TD(f_1,\mathcal{F})$*,* $\mathcal{F}_1 = \mathcal{F}_0 \backslash \{f_1\}$ *and so on.*

Define Recursive Teaching Dimension: $RTD(\mathcal{F}) = max_t TD_{min}(\mathcal{F}_t)$*, then* $RTD(\mathcal{F}) = O(VC(\mathcal{F})^2)$[2]

[2] Quadratic Upper Bound for Recursive Teaching Dimension of Finite VC Classes