CS M146 - Week 7

Xinzhu Bei

xzbei@cs.ucla.edu

February 24, 2018

Overview

- Midterm Questions
- VC Dimension
- Kernel
- Lagrangian Duality

Midterm Question: Linear Function and Hyperplane

$$y = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x} \leq 1 \\ -1 & \text{if } \mathbf{w}^T\mathbf{x} < 1 \end{cases} \tag{1}$$

Consider a linear function $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$ and a hyperplane $H : f(\mathbf{x}) = 0$

- When $b = 0$, the hyperplane H goes through the origin, while hyperplane $H' : f(\mathbf{x}) = 1$ never goes through the origin
  $\rightarrow$ When $\mathbf{x}$ and $\mathbf{w}$ are both 2-dimensional, the decision boundary does NOT go through the origin
- $-b$ is NOT the intercept. Consider 2-dimensional case:

  $$w_1x_1 + w_2x_2 + b = 0$$

  The intercept on x_1-axis is $-b/w_1$, and on x_2-axis is $-b/w_2$.
  $\rightarrow$ Without knowing $\mathbf{w}$, you cannot conclude that the hyperplane goes through some points (e.g. $(\cdot, 0)$ or $(0, \cdot)$).

Perceptron

$$y = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x} \leq 1 \\ -1 & \text{if } \mathbf{w}^T\mathbf{x} < 1 \end{cases} \tag{2}$$

Instance	1	2	3	4	5	6	7	8
Label y	+1	-1	+1	+1	+1	-1	-1	+1
Data (x_1, x_2)	(2, 0)	(2, 4)	(-1, 1)	(1, -1)	(-1, -1)	(4, 0)	(2, 2)	(0, 2)

VC Dimension

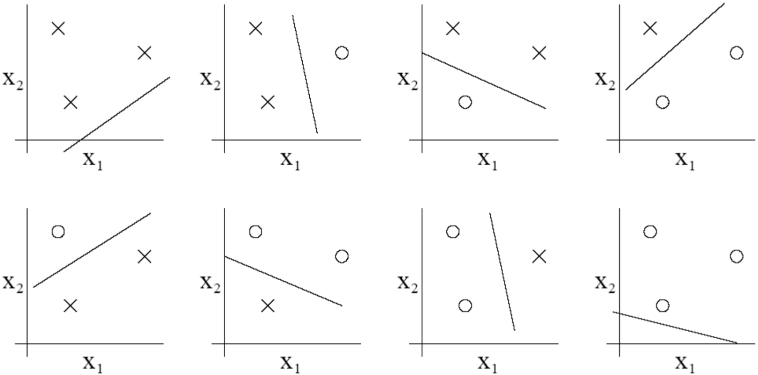
- Definition: A set S of examples is **shattered** by a set of functions H if for every partition of the examples in S into positive and negative examples there is a function in H that gives exactly these labels to the examples.
- The **VC dimension** of hypothesis space H over instance space X is the size of the largest finite subset of X that is shattered by H.
- If there exists any subset of size d that can be shattered, $VC(H) \geq d$ Even one subset will do.
- If no subset of size d can be shattered, then $VC(H) < d$

VC Dimension - Example

Consider $X = \mathbb{R}^2$, want to learn $c : X \in \{0, 1\}$

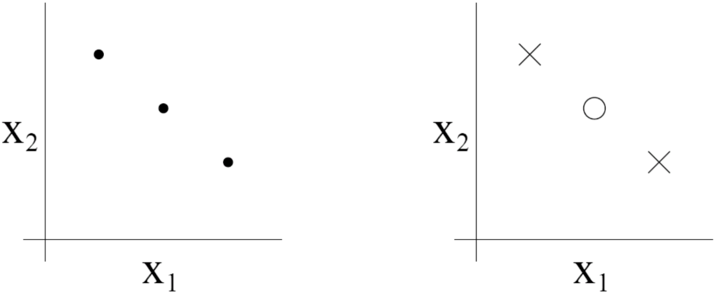
- What is VC dimension of lines in a plane?
  $H = \{((wx + b) > 0 \rightarrow y = 1)\}$

VC Dimension - Example



For any of the eight possible labelings of these points, we can find a linear classier that obtains "zero training error" on them.
Moreover, it is possible to show that there is no set of 4 points that this hypothesis class can shatter.

VC Dimension - Example



The VC dimension of H here is 3 even though there may be sets of size 3 that it cannot shatter.
Under the definition of the VC dimension, in order to prove that $VC(H)$ is at least d, we need to show only that there's at least one set of size d that H can shatter.

Kernel

Given vectors $\mathbf{x}$ and $\mathbf{z}$ in $\mathbb{R}^2$, define

$$K(\mathbf{x}, \mathbf{z}) = (1 + \mathbf{x}^T \mathbf{z})^2$$

Show that K is a kernel

- Primal SVM

$$\begin{aligned}\min_{w,b,\xi_i} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i \\ \text{s.t. } \forall i, \quad & y_i(w^T\Phi(x_i)+b) \geq 1-\xi_i \\ & \xi_i \geq 0\end{aligned} \tag{3}$$

- Dual SVM

$$\begin{aligned}\min_{\alpha} \quad & \frac{1}{2}\alpha^T Q\alpha - \mathbf{e}^T\alpha \\ \text{s.t } & 0 \leq \alpha_i \leq C, i = 1, \cdots, l \\ & \mathbf{y}^T\alpha = 0,\end{aligned} \tag{4}$$

where $Q_{ij} = y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ and $\mathbf{e} = [1, \cdots, 1]^T$.

Lagrangian Dual

Primal Problem P

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad \text{for } i = 1, \cdots, m \\
& h_i(x) = 0, \quad \text{for } i = 1, \cdots, l \\
& x \in X.
\end{aligned} \tag{5}
$$

Lagrangian Dual Problem D

$$
\begin{aligned}
\max_{u,v} \quad & \theta(u, v) \\
\text{s.t } & u \geq 0
\end{aligned} \tag{6}
$$

where

$$
\theta(u, v) = \inf_X \{ f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x) : x \in X \}
$$

is the Lagrangian dual function, and $\inf_A f = \inf\{f(x) : x \in A\}$

Lagrangian Dual

- Geometrical Interpretation of Lagrangian Dual Problem
- **Weak Duality** Theorem: Consider the primal problem P and its Lagrangian dual problem D. Let x be a feasible solution to P; that is, $x \in X$, $g(x) \geq 0$, and $h(x) = 0$. Also, let (u, v) be a feasible solution to D; that is, $u \geq 0$. Then:

$$f(x) \geq \theta(u, v)$$

** This is because

$$\begin{aligned} \theta(u, v) &= \inf\{f(\tilde{x}) + u^T g(\tilde{x}) + v^T h(\tilde{x}) : \tilde{x} \in X\} \\ &\leq f(x) + u^T g(x) + v^T h(x), \forall x \in X \leq f(x) \end{aligned} \tag{7}$$

Lagrangian Dual

Primal Problem P

$$
\begin{aligned}
\min_x \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad \text{for } i = 1, \cdots, m \\
& h_i(x) = 0, \quad \text{for } i = 1, \cdots, l \\
& x \in X.
\end{aligned} \tag{8}
$$

One circumstances of **Strong Duality**: Let X be a nonempty convex set in $\mathbb{R}^n$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ be convex, and $h\colon \mathbb{R}^n \to \mathbb{R}^l$ be affine. Suppose that the following constraint qualification is satisfied. There exists an $\hat{x} \in X$ such that $g(\hat{x}) < 0$ and $h(\hat{x}) = 0$. Then,

$$\inf\{f(x) : x \in X, g(x) \leq 0, h(x) = 0\} = \sup\{\theta(u, v) : u \geq 0\}$$

where $\theta(u, v) = \inf\{f(x) + u^T g(x) + v^T h(x) : x \in X\}$.

Deriving the Dual of SVM

Slides given from professor

The End