Lecture 5

11/11/2022

$\Rightarrow$ X $\leftarrow$ Space

$\Rightarrow$ (X, d) $\leftarrow$

$\Rightarrow$ $(X, \|\cdot\|)$

$\Rightarrow$ $(X, \langle \cdot, \cdot \rangle)$ $\longrightarrow f(\cdot)$

$\Rightarrow$ $A \in \mathbb{R}^{m \times n}$

$\Rightarrow$ EVD, SVD,

$\rightarrow$ Column space

Learning Problem

- Supervised learning

$\underline{x_i}_{J} \in \mathbb{R}^n$ $\qquad$ $\underline{Y_i}_{\ell} \in$

$\{x_i \in \mathbb{R}^n, y_i \in \mathbb{R}\}_{i=1}^{n}$ $\leftarrow$ training data

$\boxed{\{x_i, y_i\}_{i=1}^{n}} \longrightarrow \mathcal{H} \longrightarrow f(\cdot)$

$$\underline{y_i = f(x_i)}$$

x_{new} | $f(x_{new}) = y_{new}$

X : input space ; $X \subseteq \mathbb{R}^D$

Y : o/p space $Y \subseteq \mathbb{R}$

$Z = X \times Y$ called the data space

$$\boxed{y_{pred} = f_S(x_{new})}$$

R_1

x_1

Problem

⇒ Regression: $Y \subseteq \mathbb{R}$

→ Binary : $Y = \{-1, 1\}$

→ Multi-class: $Y = \{1, 2, \dots T\}$

---

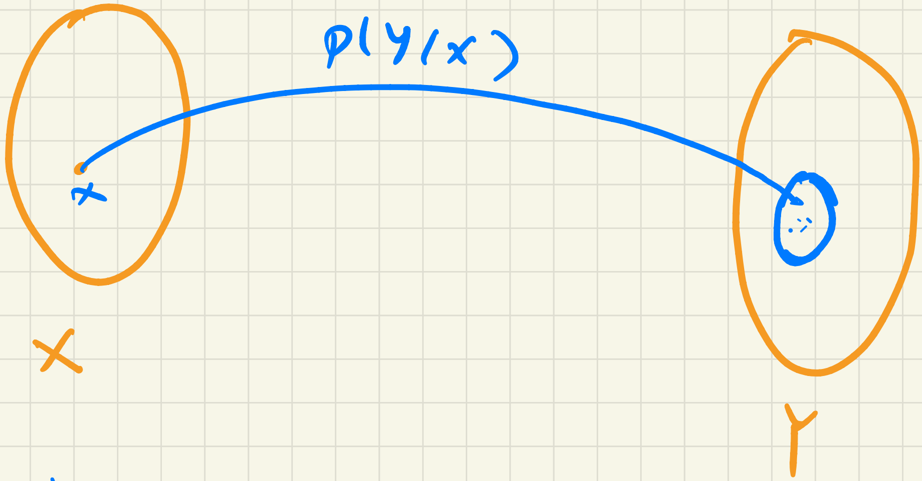
① $Z = X \times Y$, $\mu(z) = \mu(x, y)$

→ (x_i, y_i)

$$(\{x_i, y_i\}_{i=1}^n)$$

Each i/p o/p pair (x, y) is sampled from a fixed but unknown distribution

$P(x, y) = P(y|x)\,P(x)$ factorisation



$$p(x, y) = p_x(x)\, p(y|x)$$

$$\{x_i, y_i\}_{i=1}^{n} \longrightarrow \underset{\mathcal{H}}{\boxed{\quad\quad\quad}} \longrightarrow f(\cdot)$$

$\Rightarrow$ A learning algorithm is a map,

$L : Z^n \longrightarrow \mathcal{H}$ that looks at 'S'

and selects from $\mathcal{H}$ a function

$f_S : X \longrightarrow Y$ such that $f_S(x) \approx y$

in a predictive way.

x_{n+1} $\quad \{x_j, \}$ $\quad j = n+1 \ldots$

$$f(x_j) \approx y_j$$

Fn f s.t.

$f(x_i) \approx y_i \quad \forall i = 1 \ldots n$

Learning Problem

1> $(\{x_i, y_i\}) \subset Z$

2> $\mathcal{H}$

3> Loss function

4>

Loss function

A loss function $L: X \times Y \to \mathbb{R}_+$ determines the price $L(f(x), y)$, we pay if we predict $f(x)$ when in fact the true o/p is Y.

$$L(f(x), y) = (f(x) - y)^2 \Rightarrow L_2$$

$$= |f(x) - y| \Rightarrow L_1$$

$$= (|f(x) - y| - \epsilon)|_+$$

Loss function

1) x_1, y_1 : $L(f(x_1), y_1)$ ←

$\{(x_1, y_1), \ldots (x_n, y_n)\}$ → $p(x, y)$

$R\{x_i, y_i\}, f, L$

Expected Risk:

$$E[f] = \int_{X \times Y} L(f(x), y)\, p(x, y)\, dx\, dy$$

$$= \mathbb{E}_Z[L]$$

⇒ A good function should have small expected risk.

⇒ the issue is we do not know the probability $p(x, y)$

Empirical Risk

$$\mathcal{E}_n(f_n) = \frac{1}{n}\sum_{i=1}^{n} L(f(x_i), y_i)$$

↳ An estimated of the target function

$\mathcal{E}_n(f_n) \leftarrow$ Empirical Risk

$\mathbb{E}[f] \leftarrow$ Expected Risk

$$\boxed{\vdash \mathbb{E}[f] - \mathcal{E}_n(f_n) \to 0.}$$

$$\lim_{n\to\infty} \mathbb{E}_n\left[\mathcal{E}(f) - \mathcal{E}_n(f_n) \right] = 0$$

How to design an algorithm.

1) Fitting: An estimator should fit data well.

2) Stability: An estimator should be stable, it should not change much if data changes slightly

$\Rightarrow S = \{(x_i, y_i)\}_{i=1}^{n}$

$\mathcal{H}$:

$$f_S = \arg\min_{f \in \mathcal{H}} \mathcal{E}_n(f)$$

$$L(z) = (f(x) - y)^2 \quad \text{why}$$

Linear regression: Ax

$\boxed{y_i = f(x_i),\ i = 1, \ldots, n \quad f: \text{should satisfy certain additional}}$ properties.

ERM

$$\left[\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) \right]$$

Ill-posed.

$f($

Regularization

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) + \lambda \, (\text{Regul.})$$

$$\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) +$$

$$R(f) \leq A$$

Ivanov Regular

$$= \min \underbrace{\frac{1}{n}\sum_{i=1}^{n} L(f(x_i), y_i)}_{\text{goal}} + \gamma \underbrace{R(f)}_{\text{additional property}}$$

: $R(f) := \|f\|_K^2$, where $\|f\|_K^2$

is the norm in RKHS, defined by the kernel K.