1] $\hat{R}(h) = \frac{1}{n} \sum_{j=1}^{n} L(h(X_j), Y_j)$

to show: $\mathbb{P}(\lim_{n\to\infty} \hat{R}(h) = R(h)) = 1$

$$A_n(\varepsilon) = \{\omega \mid |\underbrace{\hat{R}(h)}_{= \frac{1}{n} \sum_{j=1}^{n} \underbrace{L(h(X_j), Y_j)}_{Z_i \in \{0,1\}}} - R(h)| > \varepsilon\}$$

$$\Rightarrow \mathbb{P}(|\underbrace{\overline{Z}_n}_{=\hat{R}(h)} - \overbrace{\mathbb{E}[\overline{Z}_n]}^{=R(h)}| > \varepsilon\} \leq 2e^{-2n\varepsilon^2}$$

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n(\varepsilon)) \leq 2 \sum_{n=1}^{\infty} e^{-2n\varepsilon^2} < \infty$$

Borel-Cantelli

$$\Rightarrow \quad \mathbb{P}(\lim_{n\to\infty} \hat{R}(h) = R(h)) = 1$$

$\boxed{2}$

$$R(\hat{h}) - \inf_{h\in\mathcal{H}} R(h) \leq \sup_{h\in\mathcal{H}} \underbrace{R(\hat{h}) - R(h)}_{\underbrace{\hat{R}(\hat{h}) - \hat{R}(h)}_{\leq 0} + R(\hat{h}) - \hat{R}(\hat{h}) + \hat{R}(h) - R(h)}$$

$$\leq 2 \sup_{h\in\mathcal{H}} |R(h) - \hat{R}(h)|$$

$$\mathcal{Z} = \mathcal{X} \times \mathcal{Y}, \quad z_i := (x_i, y_i), \quad z = (x, y)$$

$$f(z_1, \dots, z_n) = \underbrace{R(h) - \frac{1}{n} \sum_{i=1}^{n} \ell(h(x_i), y_i)}$$

$$|f(z_1, \dots, z_k, \dots, z_n) - f(z_1, \dots, z_k', \dots, z_n)| \leq \frac{1}{n}$$

$$\Rightarrow \mathbb{P}\Big(R(\tilde{h}) - \hat{R}(\tilde{h}) > \mathbb{E}\big[\overbrace{\sup_{h \in \mathcal{H}} R(h) - \hat{R}(h)}^{\leq 2 \mathcal{R}_n(\mathcal{L}, \mathcal{H})}\big] + t\Big)$$

$$\leq \mathbb{P}\Big(R(\tilde{h}) - \hat{R}(\tilde{h}) > \mathbb{E}[R(\tilde{h}) - \hat{R}(\tilde{h})] + t\Big)$$

$$\leq e^{-2nt^2}$$

$$\Rightarrow \mathbb{P}\left(R(h) - \hat{R}(h) > 2\mathfrak{R}_n(L, \mathcal{H}) + t \right) \leq e^{-2nt^2}$$

$$\text{analog: } \mathbb{P}\left(\hat{R}(h) - R(h) > 2\mathfrak{R}_n(L, \mathcal{H}) + t \right) \leq e^{-2nt^2}$$

$$\Rightarrow \mathbb{P}\left(\sup_{h \in \mathcal{H}} |R(h) - \hat{R}(h)| > 2\mathfrak{R}_n(L, \mathcal{H}) + t \right)$$

$$\leq \underbrace{2e^{-2nt^2}}_{=\delta} \Leftrightarrow t = \sqrt{\frac{\log(2/\delta)}{2n}}$$

3) $H \in \mathcal{H}_p$: $H = \underset{\underset{\mathbb{R}^n}{\wedge}}{a} + \operatorname{span}\{\underbrace{v_1, \dots, v_p}_{\text{orthonormal}}\}$

$$P_H : \quad \|x - P_H(x)\|^2 = \inf_{v \in H} \|x - v\|^2$$

$$= \min_{b \in \mathbb{R}^p} \|x - (a + Vb)\|^2$$

$$\Rightarrow \min_{H \in \mathcal{H}_p} \sum_{j=1}^m \|x_j - P_H(x_j)\|^2 = \min_{a, V, B} \overbrace{\sum_{j=1}^m \|x_j - (a + Vb_j)\|^2}^{= I(a, V, B)}$$

with $B = (b_1, \dots, b_m) \in \mathbb{R}^{p \times m}$

wlog $\sum_{j=1}^m b_j = 0$ $\quad V = (v_1, \dots, v_p) \in \mathbb{R}^{n \times p}$

$$\nabla_a J(a, V, B) = -2 \underbrace{\sum_{j=1}^{m} (x_j - a - V b_j)}_{} \overset{!}{=} 0$$

$$= -2 (m\bar{x} - ma)$$

$$\Rightarrow \quad a = \bar{x} = \frac{1}{m} \sum_{j=1}^{m} x_j$$

$$\nabla_{b_k} J(a, V, B) = -2 V^T (x_k - \underset{}{\overset{= \bar{x}}{\tilde{a}}} - V b_k) \overset{!}{=} 0$$

$$= \cancel{-2} \left(V^T (x_k - \bar{x}) - \underbrace{V^T V}_{= I_p} b_k \right)$$

$$b_k = V^T (x_k - \bar{x})$$

$$\rightsquigarrow \min_{v_1, \dots, v_p} \sum_{j=1}^{m} \| x_j - (\bar{x} + V V^T (x_j - \bar{x})) \|^2$$

$$= \min_{v_1 \dots v_p} \sum_{j=1}^{m} \| y_j - V V^T y_j \|^2$$

with $y_j = x_j - \bar{x}$

$$\sum_{j=1}^{m} y_j = 0$$

Lecture (PCA)

$\hookrightarrow \quad v_k = u_k$

$$\sum_{j=1}^{m} \|y_j - VV^T y_j\|^2 = tr\left((Y - VV^T Y)^T (Y - VV^T Y)\right)$$

$$\overset{V^T V = I}{=} \underbrace{tr(Y^T Y)} - tr(Y^T V V^T Y)$$

$$= tr(Y Y^T) = tr(C)$$

$$= tr(U \Lambda U^T) = tr(\Lambda)$$

$$= \sum_{i=1}^{n} \lambda_i$$

$$tr(Y^T V V^T Y) = \sum_{i=1}^{p} \langle v_i, C v_i \rangle \longrightarrow \max_{V^T V = I_p} !$$