Chapter 1 Chapter 1

- estimation ✓
- prediction ✓
- learning

(1) flip coin n times

observe $x_1, \dots, x_n$ $\quad P\{x_i = 1\} = \theta \in (0,1)$

$$\hat{\theta}(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Hoeffding $\Rightarrow$ $\boxed{P\{|\theta - \hat{\theta}| > \varepsilon\} \le \underbrace{2e^{-2n\varepsilon^2}}_{\delta}}$

$$2e^{-2n\varepsilon^2} = \delta \qquad n(\varepsilon,\delta) = \left\lceil \frac{1}{2\varepsilon^2} \ln(2/\delta) \right\rceil$$

If $n \ge n(\varepsilon,\delta)$ then $P\{|\theta - \hat{\theta}| \le \varepsilon\} \ge 1-\delta$

$\varepsilon = .01 \quad \delta = .05$

$5000 \ln 40$

---

(2) Prediction $\quad$ Suppose we know θ.

$x_1, x_2, \dots \quad Be(\theta) \qquad P(x_i = 1) = \theta$
$\qquad P(x_i = 0) = 1-\theta$

$$\boxed{f^* = \begin{cases} 1 & \text{if } \theta \ge 1/2 \\ 0 & \text{if } \theta < 1/2 \end{cases}}$$

Optimal predictor if θ is known.

$.25$ set $f^* = 0$

Loss $\quad L_\theta(f^*) = P_\theta\{x_i \ne f^*\} = \min\{\theta, 1-\theta\}$

(3) A learning algorithm maps data $X_1, \ldots, X_n$ to a predictor.

$$A_n(X_1, \ldots X_n) = \hat{f} \text{ where}$$

$$\hat{f} = \begin{cases} 0 & \text{if } \hat{\theta} \le 1/2 \\ 1 & \text{if } \hat{\theta} > 1/2 \end{cases}$$

Idea Select the predictor that works best on the training data:

$$\hat{f} = \arg\min_{f \in \{0,1\}} \Big\{ \sum_{i=1}^{n} \underbrace{\mathbb{1}_{\{X_i \neq f\}}}_{|X_i - f| \quad (X_i - f)^2} \Big\}$$