Outline for Chapter 6 : Probability

6.1 : Discrete Random Variables

6.2 : Continuous Random Variables

6.3 : Advanced topics

1. Bivariate and Multivariate normal

2. **Estimation of parameters using ML**

3. Gaussian Mixture Models and Expectation Maximisation

4. Law of Large Numbers

Parameter Estimation

$$\mathcal{P} = \{ P_\theta : \theta \in \Theta \}$$

$X_1, X_2, \dots, X_n$ drawn i.i.d from some P_θ for $\theta \in \Theta$

Estimate the true θ

Maximum Likelihood

$$x_1 \cdots x_n$$

$$\mathcal{L}(\theta) = P(X_1 = x_1, \cdots, X_n = x_n \mid \theta)$$

$$= \prod_{i=1}^{n} f_{X_i}(x_i \mid \theta)$$

$$= \prod_{i=1}^{n} p_\theta(x_i)$$

$$\log(\mathcal{L}(\theta)) = \sum_{i=1}^{n} \log(p_\theta(x_i))$$

$$R(\theta) = -\log(\mathcal{L}(\theta))$$

Bernoulli Bias

$$\mathcal{P} = \{ Bern(\theta) : \theta \in [0,1] \}$$

$$X_1, X_2, \cdots, X_n \in \{0,1\}$$

$$P_\theta(x) = \begin{cases} \theta & \text{if } x = 1 \\ 1-\theta & \text{if } x = 0 \end{cases}$$

$$= \theta^x \cdot (1-\theta)^{1-x}$$

$$R(\theta) = -\sum_{i=1}^{n} \log\left(P_\theta(x_i)\right)$$

Bernoulli Bias

$$R(\theta) = -\sum_{i=1}^{n} \log\left(p_\theta(x_i)\right)$$

$$= -\sum_{i=1}^{n} \log\left(\theta^{x_i} \cdot (1-\theta)^{1-x_i}\right)$$

$$= -\left[\sum_{i=1}^{n} x_i \log\theta + (1-x_i)\log(1-\theta)\right]$$

$$= \sum_{i=1}^{n} x_i \log\frac{1}{\theta} + (1-x_i)\log\frac{1}{1-\theta} \qquad a = \sum x_i$$

$$= a\log\frac{1}{\theta} + (n-a)\log\frac{1}{1-\theta}$$

Bernoulli Bias

$$\hat{\theta}_{ML} = \frac{a}{n}$$

$$= \frac{\sum_{i=1}^{n} x_i}{n}$$

Uniform Limits

$$\mathcal{P} = \{ \text{Unif}(a,b) : a, b \in \mathbb{R} \}$$

$$x_1, x_2, \ldots, x_n$$

$$P_\theta(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{b-a} \cdot \mathbb{1}(x \in [a,b])$$

$$R(\theta) = -\sum_{i=1}^{n} \log(P_\theta(x_i))$$

Uniform Limits

$$R(\theta) = \sum_{i=1}^{n} -\log\left(\frac{1}{b-a} \cdot \mathbb{1}(x_i \in [a,b])\right)$$

$$= \sum_{i=1}^{n} -\log\left(\frac{1}{b-a}\right) - \log\left(\mathbb{1}(x_i \in [a,b])\right)$$

If $a < \min x_i$ & $b > \max x_i$

$$R(\theta) = n \log(b-a)$$

$$a = \min x_i$$

$$b = \max x_i$$

Uniform Limits

Normal Mean

$$\mathcal{P} = \{ N(\mu, I) : \mu \in \mathbb{R} \}$$

$$x_1, x_2, \dots, x_n \qquad x_i \in \mathbb{R}$$

$$P_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (x-\mu)^2 \right)$$

$$R(\theta) = \sum_{i=1}^{n} -\log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (x_i-\mu)^2 \right) \right)$$

$$= \sum_{i=1}^{n} \frac{1}{2} (x_i - \mu)^2 + C$$

$$R'(\theta) = 0 \Rightarrow \sum_{i=1}^{n} (x_i - \mu) = 0 \Rightarrow \mu = \frac{1}{n} \sum x_i$$

Normal Mean and Variance

$$\mathcal{P} = \{ N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+ \}$$

$$x_1, x_2, \cdots, x_n \qquad x_i \in \mathbb{R}$$

$$p_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \frac{(x_i - \mu)^2}{\sigma^2} \right)$$

$$R(\theta) = -\sum_{i=1}^{n} \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2})$$

Normal Mean and Variance

$$R(\theta) = -\sum_{i=1}^{n} \left(\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2} \right)$$

$$= \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$\frac{\partial R}{\partial \mu} = \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \rightarrow 1$$

Normal Mean and Variance

$$R = \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$\frac{\partial R}{\partial \sigma^2} = \frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\cdot\frac{-1}{(\sigma^2)^2}$$

$$\frac{n}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{(\sigma^2)^2}$$

$$\sigma^2 = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n} \rightarrow \text{✓}$$

Normal Mean and Variance

Can be extended to multivariate normal

$$\mathcal{P} = \left\{ \mathcal{N}(\mu, \Sigma) : \mu \in \mathbb{R}^d, \ \Sigma \in S_d^+ \right\}$$

Data : $\{x_1, x_2, \cdots, x_N\}$ where $x_i \in \mathbb{R}^d$

ML estimates:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\Sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$$

Linear Regression With Gaussian Noise

$X \in \mathbb{R}^d, \ Y \in \mathbb{R}$

$Y = W^T X + \epsilon$ for some unknown W

$\epsilon \sim N(0, \sigma^2)$

Data = $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$

$y_i \in \mathbb{R}$

$\mathcal{P}$ for $Y|X$ = $\{N(W^T x, \sigma^2) : W \in \mathbb{R}^d\}$

Linear Regression With Gaussian Noise

$$P(Y_1 = y_1, \ldots, Y_n = y_n \mid X_1 = x_1, \ldots, X_n = x_n, w)$$

$$= \prod_{i=1}^{n} P(Y_i = y_i \mid X_i = x_i, w)$$

$$= \prod_{i=1}^{n} P(\epsilon_i = y_i - w^T x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i - w^T x_i)^2\right)$$

Linear Regression With Gaussian Noise

$$R(w) = \sum_{i=1}^{n} \frac{1}{2\sigma^2} (y_i - w^T x_i)^2 + \text{const}$$