sh3975.

Machine Learning. HW1 Shijie He

1. Academic / career goal:

Getting a job of data scientist or software engineer in company. Starting a business after years of working.

Expectation of the course:

Getting the knowledge of machine learning and being able to do some data analysis using machine learning techniques.

2.
(1) $\mathcal{L}(\theta; \{x^{(i)}\}_{i=1}^{N}) = \theta^{\sum_i x_i} (1-\theta)^{(n-\sum_i x_i)}$

$\ln(\mathcal{L}(\theta; \{x^{(i)}\}_{i=1}^{N})) = \sum_i x_i \ln\theta + (n - \sum_i x_i)\ln(1-\theta)$.

Let $\frac{\partial}{\partial\theta} \ln(\mathcal{L}) = 0$.

$\Rightarrow \sum_i x_i \frac{1}{\theta} - (n - \sum_i x_i)\frac{1}{1-\theta} = 0$

$\Rightarrow \sum x_i - \theta\sum x_i - n\theta + \theta\sum x_i = 0$

$\Rightarrow \hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$

(b) $\mathcal{L}(\theta; \{x^{(i)}\}_{i=1}^{N}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \theta_k^{x_{ik}}$

$$\ln(\mathcal{L}(\theta; \{x^{(i)}\}_{i=1}^{N})) = \sum_{i=1}^{N} \ln\left(\prod_{k=1}^{K} \theta_k^{x_{ik}}\right)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} x_{ik} \ln(\theta_k)$$

$$\ln \mathcal{L}_\lambda(\theta; \{x^{(i)}\}_{i=1}^{N}) = \ln \mathcal{L} + \lambda\left(1 - \sum_{k=1}^{K} \theta_k\right)$$

$$\frac{\partial}{\partial \theta_k} \ln \mathcal{L}_\lambda = \sum_{i=1}^{N} \frac{1}{\theta_k} x_{ik} - \lambda = 0$$

$$\Rightarrow \quad \hat{\theta}_k = \frac{1}{\lambda} \sum_{i=1}^{N} x_{ik} \quad ①$$

$$\frac{\partial}{\partial \lambda} \ln \mathcal{L}_\lambda = 1 - \sum_{k=1}^{K} \theta_k = 0$$

$$\Rightarrow \sum_{k=1}^{K} \hat{\theta}_k = 1 \quad ②$$

Plug ① to ② : $\sum_{k=1}^{K} \frac{1}{\lambda} \sum_{i=1}^{N} x_{ik} = 1$

$$\Rightarrow \frac{1}{\lambda} \sum_{i=1}^{N} \sum_{k=1}^{K} x_{ik} = 1$$

Since $\sum_{k=1}^{K} x_{ik} = 1$, $\frac{1}{\lambda} \cdot \sum_{i=1}^{N} 1 = 1$

$$\Rightarrow N = \lambda$$

Thus, $\hat{\theta}_k = \frac{1}{N} \sum_{i=1}^{N} x_{ik}$

(c). $\ln \mathcal{L}(\{\mu, \Sigma\}; \{x^{(i)}\}_{i=1}^{N})$

$$= -n\frac{d}{2}\ln(2\pi) - n\frac{1}{2}\ln(\det(\Sigma)) - \sum_{i=1}^{N}\frac{(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}{2}$$

$$\frac{\partial}{\partial\mu}(\ln\mathcal{L}) = +\sum_{i=1}^{N}\frac{1}{2}\cdot 2(x_i-\mu)\cdot\Sigma^{-1}$$

$$= \Sigma^{-1}\sum_{i=1}^{N}(x_i-\mu) = 0$$

$$\Rightarrow \hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N}x_i$$

Since $(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$ is a scalar.

$$(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = tr((x_i-\mu)^T\Sigma^{-1}(x_i-\mu))$$

$$= tr((x_i-\mu)^T(x_i-\mu)\Sigma^{-1})$$

$$\ln\mathcal{L} = n\frac{1}{2}\ln(\det(\Sigma^{-1})) - \frac{1}{2}\sum_{i=1}^{N}tr(x_i-\mu)^T(x_i-\mu)\Sigma^{-1})$$

$$\frac{\partial}{\partial\Sigma^{-1}}\ln\mathcal{L} = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}((x_i-\mu)^T(x_i-\mu))^T$$

$$= \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T(x_i-\mu) = 0.$$

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

$$\Rightarrow \hat{\Sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^{N}(x_i-\mu)^T(x_i-\mu).$$

3.

(a) Need to prove: $\int_{-\infty}^{+\infty} P(x; x_0, \Gamma)\, dx = 1$

$$\int_{-\infty}^{+\infty} \frac{1}{\pi \exp(\Gamma)\left[1 + \left(\frac{x - x_0}{\exp(\Gamma)}\right)^2\right]}\, dx$$

$$= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1 + \left(\frac{x - x_0}{\exp(\Gamma)}\right)^2}\, d\,\frac{x - x_0}{\exp(\Gamma)} \qquad \left(\text{let } \frac{x - x_0}{\exp(\Gamma)} = u\right)$$

$$= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1 + u^2}\, du$$

let $u = \tan(w) = \sin(w)/\cos(w)$.

thus, $\frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{1 + \frac{\sin^2(w)}{\cos^2(w)}} \cdot \frac{1}{\cos^2(w)}\, dw.$

$$= \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dw. = 1.$$

(b) $$E[X] = \int_{-\infty}^{+\infty} \frac{x}{\pi \exp(r)\left[1+\left(\frac{x-x_0}{\exp(r)}\right)^2\right]} dx.$$

$$= \frac{1}{\pi\exp(r)} \int_{-\infty}^{+\infty} \frac{x-x_0}{1+\left(\frac{x-x_0}{\exp(r)}\right)^2} + \frac{x_0}{1+\left(\frac{x-x_0}{\exp(r)}\right)^2} dx.$$

$$= x_0 + \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\frac{x-x_0}{\exp(r)}}{1+\left(\frac{x-x_0}{\exp(r)}\right)^2} dx.$$

let $\frac{x-x_0}{\exp(r)} = y. \Rightarrow dy = \frac{1}{\exp(r)} dx.$

$$\Rightarrow x_0 + \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{y}{1+y^2} \cdot \exp(r)\, dy.$$

$$= x_0 + \frac{\exp(r)}{\pi}\int_{-\infty}^{+\infty} \frac{1}{1+y^2} \cdot \frac{1}{2} dy^2$$

$$= x_0 + \frac{\exp(r)}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{1+y^2} dy^2.$$

We know $\int \frac{1}{1+u} du$ is indefinite.

Thus, $E[X]$ is undefined.

(c). $\ln \mathcal{L}(\{x_0, \gamma\}; \{x^{(i)}\}_{i=1}^{N})$.

$$= -N\ln(\pi) - N\gamma - \sum_{i=1}^{N} \ln\left(1 + \left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2\right)$$

$$\frac{\partial}{\partial x_0}\ln\mathcal{L} = -\sum_{i=1}^{n} \frac{1}{1+\left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2} \cdot (-2) \cdot \frac{x_i - x_0}{\exp(\gamma)} \cdot \frac{1}{\exp(\gamma)}$$

$$= \sum_{i=1}^{n} 2 \frac{\frac{x_i - x_0}{\exp(\gamma)} \cdot \frac{1}{\exp(\gamma)}}{1+\left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2}$$

$$\frac{\partial}{\partial \gamma}\ln\mathcal{L} = -\sum_{i=1}^{n} \frac{1}{1+\left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2} \cdot 2 \cdot \frac{x_i - x_0}{\exp(\gamma)} \cdot \left(-\frac{x_i - x_0}{\exp(\gamma)}\right) \cdot \exp(\gamma) - N$$

$$= \sum_{i=1}^{n} \frac{2\left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2}{1+\left(\frac{x_i - x_0}{\exp(\gamma)}\right)^2} - N.$$

4.

(a) $$\frac{\partial L}{\partial b} = -\frac{1}{N}\sum_{i=1}^{N} 2\cdot(y^{(i)} - (b + w^T x^{(i)}))$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N} 2(y^{(i)} - (b + w^T x^{(i)}))\, x^{(i)} + 2\lambda w$$