CS280 homework1_A

Name:劝琦

Student ID:2018232104

1. KL($\boldsymbol{p_{emp}}$||q)=$\int \boldsymbol{p_{emp}(x)(log p_{emp}(x) - log q(x))dx}$

$$= \int \boldsymbol{\frac{1}{n}\sum_{i=1}^{n} \delta(x, x_i)\left(log p_{emp}(x) - log q(x)\right)dx}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int \delta(x, x_i)\left(log p_{emp}(x) - log q(x)\right)dx$$
$$= \frac{1}{n}\sum_{i=1}^{n} (log p_{emp}(x_i) - log q(x_i))$$
$$= -\frac{1}{n}\sum_{i=1}^{n} log q(x_i)$$
$$\text{arg}min_q KL(p_{emp}||q) = argmax_q(-KL(p_{emp}||q))$$

$$= argmax_q(\frac{1}{n}\sum_{i=1}^{n} log q(x_i))$$

And, $\hat{\theta}$ is MLE of p(x), so about express can be as follow:

$$argmax_q \hat{\theta}$$

2.

(1)

True J(**w**) has multiple locally optimal solution .

Because lamda and –ln10/|D| …we don't know . We can't for sure J(**w**) is a convex function

(2)

$\hat{\mathbf{w}}$ is not a spare vector . Because we find all the weights of $\hat{\mathbf{w}}$ to argmin in the L2 regularized 's optimum, so almost weights of $\hat{\mathbf{w}}$ couldn't be zero .

3.

(1) $$\frac{d}{d_{\mu_k}} l(\theta) = \sum_{n=1}^{N} r_{nk} \frac{\sum_{k=1}^{K} N(x|\mu_{k,}\Sigma k)}{\sum_{k=1}^{K} \pi_k N(x|\mu_{k,}\Sigma k)}$$

(2) $$\frac{d}{d_{\Sigma_k}} l(\theta) = \sum_{n=1}^{N} r_{nk} \frac{(x_n-\mu_k)^T(x_n-\mu_k)}{2\Sigma_k}$$