Homework 2

2021年10月30日 12:54

ML2021F...
HW2 -...

HW2 - Handwritten Assignment

1.

Consider a generative classification model for K classes defined by prior class probabilities $p(C_k) = \pi_k$ and general class-conditional densities $p(x|C_k)$, where x is the input feature vector. (Note that $\pi_1 + \ldots + \pi_k = 1$)
Suppose we are given a training data set $\{x_n, \mathbf{t_n}\}$ where $n = 1, \ldots, N$,and $\mathbf{t_n}$ is a binary target vector of length K that uses the $1 - of - K$ coding scheme, so that it has components $t_{nk} = 1$ if pattern n is from class C_k, otherwise $t_{nj} = 0$. Assuming that the data points are drawn independently from this model, show that the maximum-likelihood solution for the prior problablities is given by

$$\pi_k = \frac{N_k}{N}$$

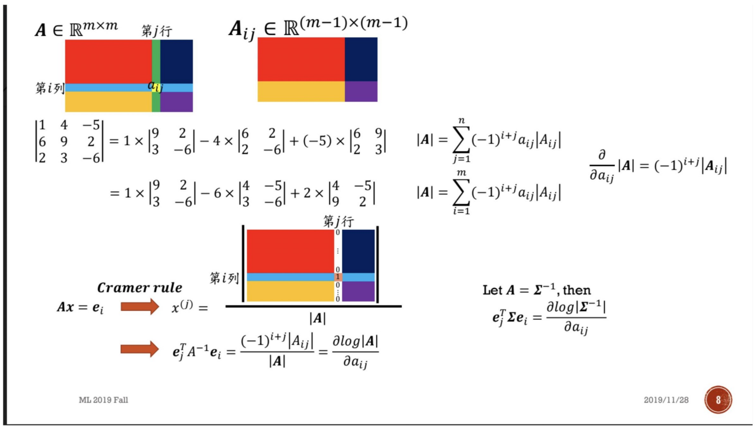
where N_k is the number of data points assigned to class C_k.

2.

Show that

$$\frac{\partial log(det\ \mathbf{\Sigma})}{\partial \sigma_{ij}} = \mathbf{e_j} \mathbf{\Sigma}^{-1}\ \mathbf{e_i^T}$$

where $\mathbf{\Sigma} \in \mathbb{R}^{m \times m}$ is a (non-singular) covariance matrix and $\mathbf{e_j}$ is a row vector(ex: $e_3 = [0, 0, 1, 0, \ldots, 0]$).
Hint:



3.

Consider the classification model of $\mathbf{problem\ 1}$ & result of $\mathbf{problem\ 2}$ and now suppose that the class-condition densities are given by Gaussian distributions with a shared convariance matrix, so that

$$p(x|C_k) = \mathcal{N}(x|\mu_\mathbf{k}, \mathbf{\Sigma})$$

Show that the maximum likelihood solution for the mean of the Gaussian distribution for class C_k is given by

$$\mu_{\mathbf{k}} = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}x_n$$

which represents the mean of those feature vectors assigned to class C_k. Similarly, show that the maximum likelihood solution for the shared covariance matrix is given by

$$\mathbf{\Sigma} = \sum_{k=1}^{K}\frac{N_k}{N}\mathbf{S_k}$$

where

$$\mathbf{S_k} = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(\mathbf{x_n} - \mu_{\mathbf{k}})(\mathbf{x_n} - \mu_{\mathbf{k}})^T$$

Thus $\mathbf{\Sigma}$ is given by a weighted average of the covariance of the data associated with each class, in which the weighting coeffients are given by the prior probabilities of the classes.

HW2 - Handwritten Assignment

1.

Consider a generative classification model for K classes defined by prior class probabilities $p(C_k) = \pi_k$ and general class-conditional densities $p(x|C_k)$, where x is the input feature vector. (Note that $\pi_1 + \ldots + \pi_k = 1$)
Suppose we are given a training data set $\{x_n, \mathbf{t_n}\}$ where $n = 1, \ldots, N$,and $\mathbf{t_n}$ is a binary target vector of length K that uses the $1 - of - K$ coding scheme, so that it has components $t_{nk} = 1$ if pattern n is from class C_k, otherwise $t_{nj} = 0$. Assuming that the data points are drawn independently from this model, show that the maximum-likelihood solution for the prior probablities is given by

$$\pi_k = \frac{N_k}{N}$$

where N_k is the number of data points assigned to class C_k.

Maximize log likelihood s.t. $\sum_{k=1}^{K} \pi_k = 1$

Using Langrange Multiplier,

$$L(\pi,\lambda) = \sum_{n=1}^{N}\sum_{k=1}^{K} y_{n,k}\left[\log(P(x_n|C_k) + \log\pi_k\right] + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

To maximize the above equation, we take $\nabla_{\pi_k} L(\pi,k) = 0$:

$$\frac{\partial}{\partial \pi_k} L(\pi,\lambda) = \frac{1}{\pi_k}\sum_{n=1}^{N} y_{n,k} + \lambda = 0$$

$$\sum_{n=1}^{N} y_{n,k} = -\lambda\pi_k$$

$$\sum_{n=1}^{N} y_{n,k} = -\lambda \pi_k$$

$$\pi_k = -\frac{\sum_{n=1}^{N} y_{n,k}}{\lambda}$$

$$\pi_k = -\frac{N_k}{\lambda} \quad -(*)$$

Taking $\nabla_\lambda L(\pi, \lambda) = 0$,

$$\frac{\partial}{\partial \lambda} L(\pi, \lambda) = \sum_{k=1}^{K} \pi_k - 1 = 0$$

$$\sum_{k=1}^{K} \pi_k = 1$$

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \frac{-N_k}{\lambda}$$

$$\sum_{k=1}^{K} \pi_k = -\frac{N}{\lambda} = 1$$

$$\lambda = -N$$

From $(*)$,

$\Rightarrow$ $\pi_k = \frac{N_k}{N}$.

2. Show that

$$\frac{\partial log(det\ \mathbf{\Sigma})}{\partial \sigma_{ij}} = \mathbf{e_j} \mathbf{\Sigma}^{-1} \mathbf{e_i^T}$$

where $\mathbf{\Sigma} \in \mathbb{R}^{m \times m}$ is a (non-singular) covariance matrix and $\mathbf{e_j}$ is a row vector(ex: $e_3 = [0, 0, 1, 0, \ldots, 0]$).

Hint:

$$\frac{\partial}{\partial \sigma_{ij}} \log|\Sigma| = \frac{1}{|\Sigma|} \frac{\partial}{\partial \sigma_{ij}} |\Sigma|$$

$$= \frac{(-1)^{i+j} |A_{ij}|}{|\Sigma|}$$

$$= e_j \Sigma^{-1} e_i^T$$

Hence proved.

3. Consider the classification model of **problem 1** & result of **problem 2** and now suppose that the class-condition densities are given by Gaussian distributions with a shared convariance matrix, so that

$$p(x|C_k) = \mathcal{N}(x|\mu_k, \Sigma)$$

Show that the maximum likelihood solution for the mean of the Gaussian distribution for class C_k is given by

$$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk} x_n$$

which represents the mean of those feature vectors assigned to class C_k. Similarly, show that

$$p(x|C_k) = N(x|\mu_k, \Sigma)$$

which represents the mean of those feature vectors assigned to class C_k. Similarly, show that the maximum likelihood solution for the shared covariance matrix is given by

$$\Sigma = \sum_{k=1}^{K} \frac{N_k}{N} \mathbf{S_k}$$

where

$$\mathbf{S_k} = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(\mathbf{x_n} - \mu_k)(\mathbf{x_n} - \mu_k)^T$$

Thus Σ is given by a weighted average of the covariance of the data associated with each class, in which the weighting coeffients are given by the prior probabilities of the classes.

Likelihood $= p(t, x | \pi, \mu, \Sigma) = \prod_{n=1}^{N} \prod_{j=1}^{K} [\pi_j N(x|\mu_k, \Sigma)]^{t_{nj}}$

Taking log:

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{K} t_{nj}[K + \ln|\Sigma| + (x_n-\mu_j)^T \Sigma^{-1}(x_n-\mu_j)] \quad -(*)$$

Taking $\nabla_{\mu_k}(*) = 0$, we have

$$\sum_{n=1}^{K} t_{nk} \Sigma^{-1}(x_n - \mu_k) = 0$$

$$N_k \Sigma^{-1}(x - \mu_k) = 0$$

$$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk} x_n$$

Next, From $(*)$, the terms dependent on Σ are:

$$\frac{N}{} \ln|\Sigma| - \frac{1}{2}\sum^{K} \sum (x_n - \mu_j)^T \Sigma^{-1}(x_n - \mu_j)$$

$$-\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{j=1}^{K}\sum_{n\in C_j}(x_n-\mu_j)^T\Sigma^{-1}(x_n-\mu_j)$$

$$= -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{j=1}^{K}\sum_{n\in C_j}\operatorname{tr}\left(\Sigma^{-1}(x_n-\mu_j)(x_n-\mu_j)^T\right)$$

$$= -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\sum_{j=1}^{K}\sum_{n\in C_j}\Sigma^{-1}(x_n-\mu_j)(x_n-\mu_j)^T\right)$$

$$= -\frac{N}{2}\ln|\Sigma| - \frac{N}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{1}{N}\sum_{j=1}^{K}\sum_{n\in C_j}(x_n-\mu_j)(x_n-\mu_j)^T\right)$$

$$= -\frac{N}{2}\ln|\Sigma| - \frac{N}{2}\operatorname{tr}\left(\Sigma^{-1}S\right) \quad -(*)$$

Taking $\nabla_\Sigma (*) = 0$, we have

$$\Sigma = \sum_{k=1}^{K}\frac{N_k}{N}S_k.$$

Hence proved.