第三章作业一 计算：

1. 如下所示，随机产生 1000 个样本；

a. Generate and plot a data set of $N = 1{,}000$ two-dimensional vectors that stem from three equiprobable classes modeled by normal distributions with mean vectors $m_1 = [1,\ 1]^T$, $m_2 = [7,\ 7]^T$, $m_3 = [15,\ 1]^T$ and covariance matrices $S_1 = \begin{bmatrix} 12 & 0 \\ 0 & 1 \end{bmatrix}$, $S_2 = \begin{bmatrix} 8 & 3 \\ 3 & 2 \end{bmatrix}$, $S_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

b. Repeat (a) when the *a priori* probabilities of the classes are given by the vector $P = [0.6,\ 0.3,\ 0.1]^T$.

2. 基于上述样本集，分别采用 MLE 和 BE 的方法，估计三类的参数；
3. 从上述样本集中，随机取 300 个样本，重复上述实验；
4. 分析三次实验的参数估计精度，分析误差产生的原因。

第三章作业二 推导：

17. The purpose of this problem is to derive the Bayesian classifier for the d-dimensional multivariate Bernoulli case. As usual, work with each class separately, interpreting $P(\mathbf{x}|\mathcal{D})$ to mean $P(\mathbf{x}|\mathcal{D}_i, \omega_i)$. Let the conditional probability for a given category be given by

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

and let $\mathcal{D} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be a set of n samples independently drawn according to this probability density.

(a) If $\mathbf{s} = (s_1, ..., s_d)^t$ is the sum of the n samples, show that

$$P(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(b) Assuming a uniform a priori distribution for $\boldsymbol{\theta}$ and using the identity

$$\int_0^1 \theta^m (1-\theta)^n \, d\theta = \frac{m!n!}{(m+n+1)!},$$

show that

$$p(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

(c) Plot this density for the case $d = 1, n = 1$, and for the two resulting possibilities for s_1.

(d) Integrate the product $P(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})$ over $\boldsymbol{\theta}$ to obtain the desired conditional probability

$$P(\mathbf{x}|\mathcal{D}) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i}\left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}.$$

(e) If we think of obtaining $P(\mathbf{x}|\mathcal{D})$ by substituting an estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in $P(\mathbf{x}|\boldsymbol{\theta})$, what is the effective Bayesian estimate for $\boldsymbol{\theta}$?

第三章作业三 HMM：

11. Consider the use of hidden Markov models for classifying sequences of four visible states, A–D. Train two hidden Markov models, each consisting of three hidden states (plus a null initial state and a null final state), fully connected, with the following data. Assume that each sequence starts with a null symbol and ends with an end null symbol (not listed).

sample	ω_1	ω_2
1	AABBCCDD	DDCCBBAA
2	ABBCBBDD	DDABCBA
3	ACBCBCD	CDCDCBABA
4	AD	DDBBA
5	ACBCBABCDD	DADACBBAA
6	BABAADDD	CDDCCBA
7	BABCDCC	BDDBCAAAA
8	ABDBBCCDD	BBABBDDDCD
9	ABAAACDCCD	DDADDBCAA
10	ABD	DDCAAA

(a) Print out the full transition matrices for each of the models.

(b) Assume equal prior probabilities for the two models and classify each of the following sequences: ABBBCDDD, DADBCBAA, CDCBABA, and ADBBBCD.

(c) As above, classify the test pattern BADBDCBA. Find the prior probabilities for your two trained models that would lead to equal posteriors for your two categories when applied to this pattern.