Imperial College
London

COURSEWORK 1

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Mathematics for Machine Learning

Author:
Arnold Cheung (CID: 01184493)

Date: October 15, 2019

1 Statistics and Probabilities

Question 1

Given the dataset $\mathcal{D}$:

$$\mathcal{D} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix} \right\} \tag{1}$$

The sample mean $\overline{\boldsymbol{\mathcal{D}}}$ is given by a column vector whose jth element is equal to the mean of the jth element in each of the vectors in the dataset, where N is the number of samples in the dataset $\mathcal{D}$:

$$\overline{\mathcal{D}}_j = \frac{1}{N} \sum_{i=1}^{N} \mathcal{D}_{ij} \tag{2}$$

$$\overline{\boldsymbol{\mathcal{D}}} = \frac{1}{N} \sum_{i=1}^{N} \overline{\boldsymbol{\mathcal{D}_i}} \tag{3}$$

For $N = 3$, the sample mean $\overline{\boldsymbol{\mathcal{D}}}$ is therefore:

$$\overline{\boldsymbol{\mathcal{D}}} = \frac{1}{3} \begin{bmatrix} 1-1-4 \\ 2+0+4 \\ 3+0+2 \end{bmatrix} = \begin{bmatrix} \frac{-4}{3} \\ 2 \\ \frac{5}{3} \end{bmatrix} \tag{4}$$

The sample covariance matrix using $1/N$ (biased estimate) $Cov(\mathcal{D})$ is computed as followswhere $\overline{\boldsymbol{\mathcal{D}}}$ is the sample mean column matrix and $\boldsymbol{\mathcal{D}_i}$ is the ith sample in the dataset $\mathcal{D}$:

$$Cov(\mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{\mathcal{D}_i} - \overline{\boldsymbol{\mathcal{D}}})(\boldsymbol{\mathcal{D}_i} - \overline{\boldsymbol{\mathcal{D}}})^T \tag{5}$$

$Cov(\mathcal{D})$ is a KxK matrix where K is the number of entries in each sample column vector:

$$Cov(\mathcal{D}) = \frac{1}{3} \left[\frac{1}{9} \begin{bmatrix} 49 & 0 & 28 \\ 0 & 0 & 0 \\ 28 & 0 & 16 \end{bmatrix} + \frac{1}{9} \begin{bmatrix} 1 & -6 & -5 \\ -6 & 36 & 30 \\ -5 & 30 & 25 \end{bmatrix} + \frac{1}{9} \begin{bmatrix} 64 & -48 & -8 \\ -48 & 36 & -6 \\ -8 & -6 & 1 \end{bmatrix} \right] \tag{6}$$

$$= \begin{bmatrix} \frac{38}{9} & -2 & \frac{5}{9} \\ -2 & \frac{8}{3} & \frac{4}{3} \\ \frac{5}{9} & \frac{4}{3} & \frac{14}{9} \end{bmatrix} \tag{7}$$

Question 2

Below shows two sets of data, each with 100 data points with mean

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

and marginal variances $\sigma_1^2 = 2$ and $\sigma_2^2 = 0.5$.

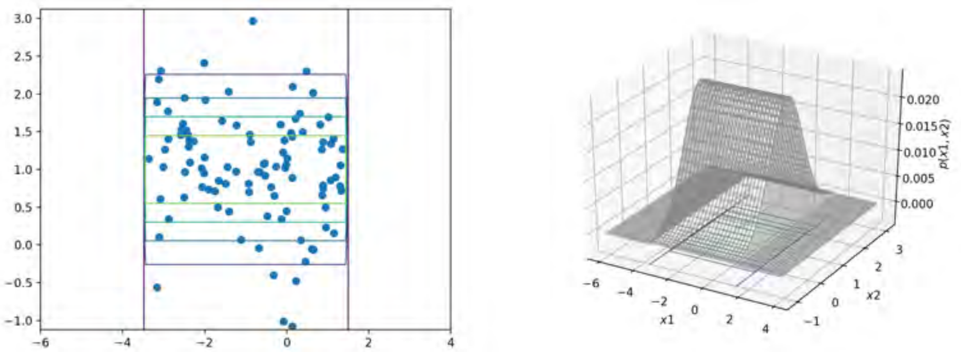

Figure 1: Dataset 1, A joint distribution of a uniform distribution with mean $= -1$ and variance $= 2$ and a Gaussian distribution with mean $= 1$ and variance $= 0.5$

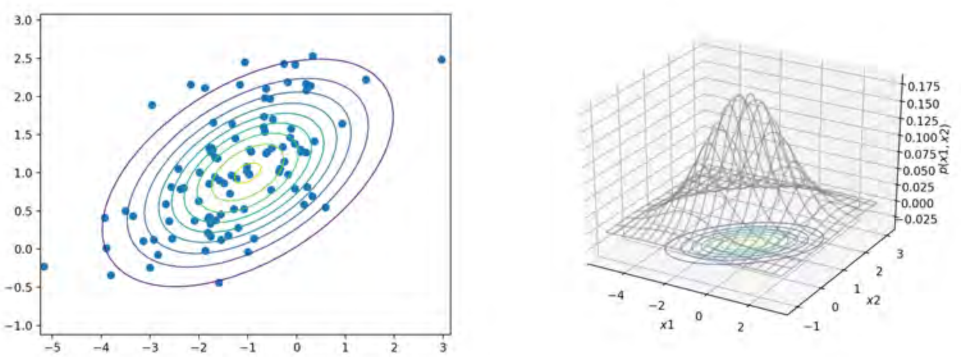

Figure 2: Dataset 2, A bivariate Gaussian distribution with mean $= [-1, 1]$ and marginal variances$= [2, 0.5]$

Both of the datasets have the same mean, $\boldsymbol{\mu}$ and marginal variances σ_1^2 and σ_2^2. However, the shape of the two distributions are clearly different as shown in Figure 1 and 2. The shapes of the distributions have been ensured to be different. The two entries of data points in dataset 1 have been generated from a different one-dimensional distribution, where x_1 is chosen from a normal distribution:

$$\mathcal{U}(-\frac{\sqrt{24}}{2} - 1, \frac{\sqrt{24}}{2} - 1) \tag{8}$$

and x_2 is chosen from a one-dimensional Gaussian distribution:

$$\mathcal{N}(1, 0.5) \tag{9}$$

While dataset 2 is generated from a bivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and the covariance matrix

$$\boldsymbol{Cov}(D) = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2 \end{bmatrix} \tag{10}$$

The shape of the distribution is tilted due to the covariance between the univariate Gaussian distribution that forms the bivariate Gaussian distribution.

Question 3

The probability of successful compilation is described by the Bernoulli distribution:

$$P(x|\mu) = \mu^x(1-\mu)^{1-x} \tag{11}$$

The likelihood is given by a binomial distribution with mean μ and the number of trials $N = 20$:

$$P(x|\mu, 20) = \binom{20}{x}\mu^x(1-\mu)^{20-x} \tag{12}$$

The conjugate prior beta distribution with parameters $\alpha = 2$ and $\beta = 2$ is given by:

$$P(\mu) = \mu^{\alpha-1}(1-\mu)^{\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{13}$$

$$\tag{14}$$

The joint distribution is given by:

$$P(x|\mu, 20)P(\mu) = \binom{20}{x}\mu^x(1-\mu)^{20-x}\mu^{\alpha-1}(1-\mu)^{\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{15}$$

$$= \binom{20}{x}\mu^{x+\alpha-1}(1-\mu)^{20-x+\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{16}$$

The evidence $P(x)$ is the marginal distribution of x:

$$P(x) = \int_0^1 \binom{20}{x}\mu^{x+\alpha-1}(1-\mu)^{20-x+\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}d\mu \tag{17}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{20}{x}\frac{\Gamma(\alpha+x)\Gamma(\beta+20-x)}{\Gamma(\alpha+\beta+20)} \tag{18}$$

For successes $x = 6$, $\alpha = 2$ and $\beta = 2$, the posteria distribution on μ is therefore:

$$\frac{P(x|\mu, 20)P(\mu)}{P(x)} = \frac{\Gamma(\alpha+\beta+20)}{\Gamma(\alpha+x)\Gamma(\beta+20-x)}\mu^{x+\alpha-1}(1-\mu)^{20-x+\beta-1} \tag{19}$$

$$= \frac{\Gamma(24)}{\Gamma(8)\Gamma(16)}\mu^7(1-\mu)^{15} \tag{20}$$

$$= 3922511\mu^7(1-\mu)^{15} \tag{21}$$

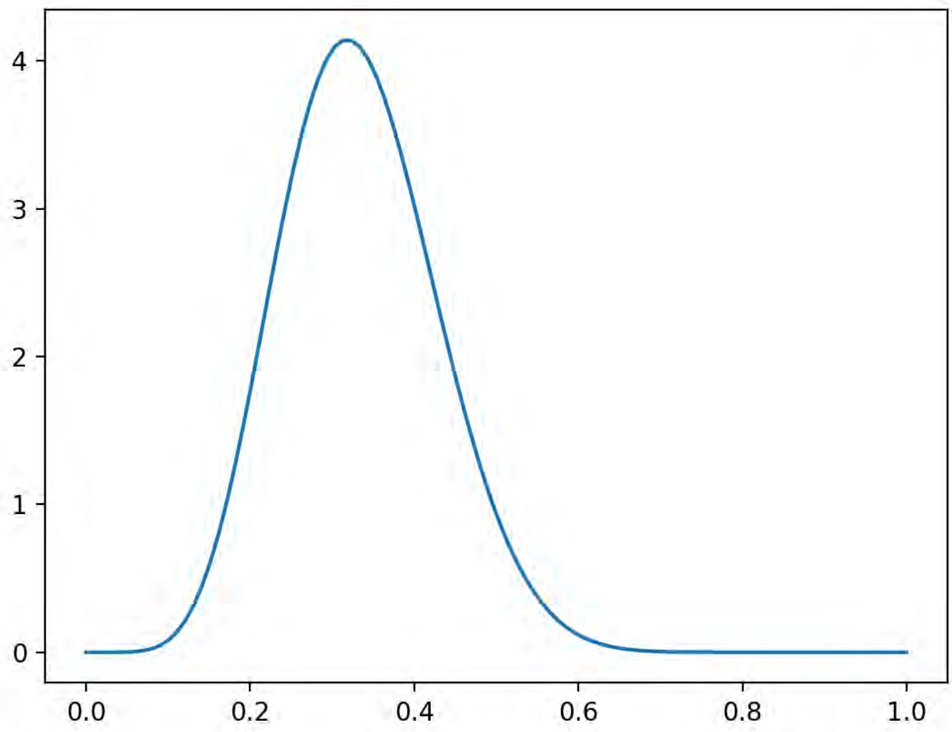

Figure 3: The plot of the posterior distribution on μ

Both the prior and the posterior are beta distributions, as the beta distribution is a conjugate prior of the Bernoulli distribution and the binomial distribution. The hyperparameters α and β have however changed from $\alpha = 2$ and $\beta = 2$ to $\alpha' = \alpha + x = 8$ and $\beta' = \beta + N - x = 16$, where α' and β' are the new hyperparameters of the posterior beta distribution, $N = 20$ is the number of trials and $x = 6$ is the number of observed success.

2 Graphical Models

Question 1

$$p(a,b,c,d,e,f) = p(a|b,c)p(c|b)p(d)p(e|d,a)p(f|c,d,e) \tag{22}$$

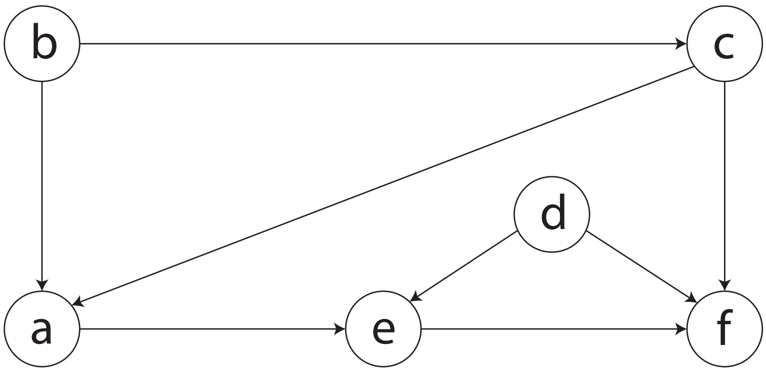

Figure 4: The graphical model of $p(a,b,c,d,e,f)$

Question 2

(a) The statement $a \perp\!\!\!\perp g|k$ is true, all paths from a to g are blocked by e as e is not in the set $\{k\}$.

(b) The statement $d \perp\!\!\!\perp h|i$ is false, one unblocked path is $d \rightarrow j \rightarrow h$.

(c) The statement $g \perp\!\!\!\perp a|c$ is true, all paths from g to a are blocked by e as e is not in the set $\{c\}$.

(d) The statement $e \perp\!\!\!\perp j|k$ is false, one unblocked path is $e \rightarrow d \rightarrow j$.

(e) The statement $v \perp\!\!\!\perp e|j$ is false, one unblocked path is $b \rightarrow d \rightarrow e$.

(f) The statement $j \perp\!\!\!\perp c|\{k, g\}$ is false, one unblocked path is $j \rightarrow h \rightarrow i \rightarrow c$.

(g) The statement $a \perp\!\!\!\perp k$ is false, one unblocked path is $a \rightarrow b \rightarrow d \rightarrow k$.

(h) The statement $a \perp\!\!\!\perp k|e$ is false, one unblocked path is $a \rightarrow b \rightarrow d \rightarrow k$.

(i) The statement $h \perp\!\!\!\perp d|\{j, e\}$ is false, one unblocked path is $h \rightarrow i \rightarrow d$.

(j) The statement $b \perp\!\!\!\perp h|e$ is false, one unblocked path is $b \rightarrow j \rightarrow h$, as e is a descendant of j which is in the set $\{e\}$.

(k) The statement $h \perp\!\!\!\perp c|d$ is false, one unblocked path is $h \rightarrow i \rightarrow c$.

(l) The statement $a \perp\!\!\!\perp f|k$ is false, one unblocked path is $a \rightarrow b \rightarrow j \rightarrow f$, as k is a descendant of j which is in the set $\{k\}$.

(m) The statement $i \perp\!\!\!\perp a|j$ is false, one unblocked path is $i \rightarrow h \rightarrow j \rightarrow b \rightarrow a$, as j is in the set $\{j\}$.

(n) The statement $g \perp\!\!\!\perp h|e$ is false, one unblocked path is $g \rightarrow e \rightarrow d \rightarrow j \rightarrow h$, as e is in the set $\{e\}$.

(o) The statement $g \perp\!\!\!\perp f|d$ is true, all paths from g to f are blocked by e as e is not in the set $\{d\}$.

(p) The statement $g \perp\!\!\!\perp h$ is true, all paths from g to h are blocked by e as e is not in the set $\emptyset$.

(q) The statement $a \perp\!\!\!\perp i|g$ is true, all paths from a to i are blocked by j and d as neither has descendants in the set $\{g\}$.

(r) The statement $b \perp\!\!\!\perp h$ is true, all paths from b to h are blocked by j and d as neither has descendants in the set $\emptyset$.

(s) The statement $b \perp\!\!\!\perp h|g$ is true, all paths from b to h are blocked by j and d as neither has descendants in the set $\{g\}$.

(t) The statement $i \perp\!\!\!\perp a$ is true, all paths from i to a are blocked by j and d as neither has descendants in the set $\emptyset$.