CK0146 or TIP8311: Homework 02

Exercise 02.01 ✓(2.2 PRML). The form of the Bernoulli distribution given by $\text{Bern}(x|\mu) = \mu^x(1-\mu)^{1-\mu}$ is not symmetric between the two values of x. In some situations, it will be more convenient to use an equivalent formulation for which $x \in \{-1, +1\}$, in which case the distribution can be written

$$p(x|\mu) = \Big(\frac{1-\mu}{2}\Big)^{(1-x)/2}\Big(\frac{1+\mu}{2}\Big)^{(1+x)/2}, \quad \text{where } \mu \in [-1, +1]. \tag{1}$$

Show that the distribution above is normalised and evaluate its mean, variance and entropy.

Exercise 02.02 ✓(2.13 PRML). Evaluate the Kullback-Leibler divergence

$$\mathbb{KL}[p|q] = -\int p(\mathbf{x})\ln q(\mathbf{x})\mathrm{d}\mathbf{x} - \Big[-\int p(\mathbf{x})\ln p(\mathbf{x})\mathrm{d}\mathbf{x}\Big] = -\int p(\mathbf{x})\ln\Big[\frac{q(\mathbf{x})}{p(\mathbf{x})}\Big]\mathrm{d}\mathbf{x}.$$

between two Gaussians $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{m}, \mathbf{L})$.

Exercise 02.03 ✓(2.15 PRML). Show that the entropy of the multivariate Gaussian $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$\mathbb{H}[\mathbf{x}] = \frac{1}{2}\ln|\boldsymbol{\Sigma}| + \frac{D}{2}(1 + \ln(2\pi)), \tag{2}$$

where D is the dimensionality of $\mathbf{x}$.

Exercise 02.04 ✓(2.19 PRML). Show that a real symmetric matrix $\boldsymbol{\Sigma}$ having the eigenvector equation $\boldsymbol{\Sigma}\mathbf{u}_i = \lambda_i\mathbf{u}_i$, where $i = 1\dots, D$, can be expressed as an expansion in the eigenvectors, with coefficients given by the eigenvalues, of the form $\boldsymbol{\Sigma} = \sum_{i=1}^{D}\lambda_i\mathbf{u}_i\mathbf{u}_i^T$.

Similarly, show that the inverse matrix $\boldsymbol{\Sigma}^{-1}$ has a representation of the form $\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$.

Exercise 02.05 ✓(2.21 PRML). Show that a real symmetric matrix of size $D \times D$ has $D(D+1)/2$ independent parameters.

Exercise 02.06 ✓. Let $\boldsymbol{\Sigma}^{(F)}$, $\boldsymbol{\Sigma}^{(D)}$ and $\boldsymbol{\Sigma}^{(I)}$ be three randomly generated $M \times M$ covariance matrixes of a Gaussian distribution. Let $\boldsymbol{\Sigma}^{(F)}$ be of general form (F, for full), $\boldsymbol{\Sigma}^{(D)}$ be diagonal (D) and $\boldsymbol{\Sigma}^{(I)}$ be proportional to an identity matrix (I, for isotropic). In, addition, let $\boldsymbol{\mu}^{(F)} = \boldsymbol{\mu}^{(D)} = \boldsymbol{\mu}^{(I)} \neq \mathbf{0}$ be the corresponding random mean vectors.

For $M = 2$,

- plot the contours of constant density for the three Gaussian distributions;
- show graphically (and atop of the iso-density contours) all of the eigenpairs $\{(\mathbf{u}_i^{(j)}, \lambda_i^{(j)}) : \forall i\}_{j\in\{F,D,I\}}$ of eigenvectors and associated eigenvalues of the respective covariance matrices.

For $M = 5$,

- evaluate the expressions for the mean and the (co)variance of the conditional distributions $p(x_1^{(j)}|\mathbf{x}_{2:D}^{(j)})$ and $p(\mathbf{x}_{1:2}^{(j)}|\mathbf{x}_{3:D}^{(j)})$ with $j \in \{F, D, I\}$, and plot their iso-density contours. For conditioning, let $\mathbf{x}_{2:D}^{(j)}$ and $\mathbf{x}_{3:D}^{(j)}$ be two random vectors;
- evaluate the expressions for the mean and the (co)variance of the marginal distributions $p(x_1^{(j)})$ and $p(\mathbf{x}_{1:2}^{(j)})$ again with $j \in \{F, D, I\}$, and plot their iso-density contours.

Exercise 02.07 ✓. You are given (download me here) a dataset $\mathcal{X} = \{(\mathbf{x}(n), c(n))\}_{n=1}^{N}$ consisting of N observations of some unknown 2D probability distribution $p(\mathbf{x})$ and associated class label c.

You are requested to estimate and visualise:

- the mono-dimensional probability distributions $p(x_1)$ and $p(x_2)$ using the i) histogram method (consider several values of the bin width Δ and qualitatively select a ''visually optimal' one); and the ii) K nearest neighbour method (again, consider several values of K and qualitatively select a 'visually optimal' one);

- the bi-dimensional probability distribution $p(\mathbf{x})$ using the i) K nearest neighbour method and ii) the kernel method with a Gaussian function. Again, for both estimators, consider several values of K and h, respectively, and qualitatively select a 'visually optimal' one.

You are also requested to build a K-NN classifier using the first 80% of the observations as *learning* set and the remaining 20% for validation. Consider several values of K and select the K associated with the best classification performance (use overall misclassification rate) on the validation set.