FMAN-45 Machine Learning, Fall 2016
Assignment 2

Solve the problems and write down the solutions. If the assignment involves programming, download the code we provide and do the additional, required programming. Write a detailed report. All solutions, plots and figures should be in one *pdf. It should be possible to understand all material presented in the report without running any code.* Submit your solutions and code using your individual Moodle account *as two files (a pdf and a single archive with all the code) at* `http://moodle.maths.lth.se/course/` *by the deadline (note that there will be no extensions).*

1 The Gaussian distribution's normalization (50 points)

In this exercise, we prove the normalization condition,

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1,$$

for the univariate Gaussian. To do this consider, the integral

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx$$

which we can evaluate by first writing its square in the form

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}y^2\right) dxdy.$$

Now make the transformation from Cartesian coordinates (x, y) to polar coordinates (r, θ) and then substitute $u = r^2$.

(a) (25 points) Show that, by performing the integrals over θ and u, and then taking the square root of both sides, we obtain

$$I = \left(2\pi\sigma^2\right)^{1/2}.$$

(b) (25 points) Finally, use this result to show that the Gaussian distribution $\mathcal{N}(x|\mu, \sigma^2)$ is normalized.

2 Hypercubes (50 points)

Consider a sphere of radius a in D-dimensions together with the concentric hypercube of side $2a$, so that the sphere touches the hypercube at the centres of each of its sides. The surface area of a unit D-dimensional sphere is

$$S_D = \frac{2\pi^{D/2}}{\Gamma(D/2)}.$$

The volume of a D-dimensional sphere with radius r is $V_D = \frac{S_D r^D}{D}$.

(a) (25 points) Use the above results to show that the ratio of the volume of the sphere to the volume of the cube is given by

$$\frac{\text{volume of sphere}}{\text{volume of cube}} = \frac{\pi^{D/2}}{D 2^{D-1} \Gamma(D/2)} \tag{1}$$

Now make use of Stirling's formula in the form

$$\Gamma(x+1) \simeq (2\pi)^{1/2} e^{-x} x^{x+1/2}$$

which is valid for $x \gg 1$, to show that, as $D \to \infty$, the ratio (1) goes to zero.

(b) (25 points) Show also that the ratio of the distance from the centre of the hypercube to one of the corners, divided by the perpendicular distance to one of the sides, is $\sqrt{D}$, which therefore goes to ∞ as $D \to \infty$.

From these results we see that, in a space of high dimensionality, most of the volume of a cube is concentrated in the large number of corners, which themselves become very long spikes!

3 The Dirichlet distribution's normalization (50 points)

The Dirichlet distribution is given by

$$\mathrm{Dir}(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1},$$

where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$. It also holds that $\mu_k \in [0,1]$ and $\sum_{k=1}^{M} \mu_k = 1$. In this exercise, we prove the normalization of the Dirichlet distribution using induction. Use the fact that the beta distribution, which is a special case of the Dirichlet for $M = 2$, is normalized (you do not need to prove this). We now assume that the Dirichlet distribution is normalized for $M-1$ variables and prove that it is normalized for M variables. To do this, consider the Dirichlet distribution over M variables, and take account of the constraint $\sum_{k=1}^{M} \mu_k = 1$ by eliminating μ_M, so that the Dirichlet is written

$$p_m(\mu_1, \ldots, \mu_{M-1}) = C_M \prod_{k=1}^{M-1} \mu_k^{\alpha_k - 1} \left(1 - \sum_{j=1}^{M-1} \mu_j\right)^{\alpha_M - 1}$$

and our goal is to find an expression for C_M. To do this, integrate over μ_{M-1}, taking care over the limits of integration, and then make a change of variable so that this integral has limits 0 and 1. By assuming the correct result for C_{M-1} and making use of

$$\int_0^1 \mu^{a-1} (1-\mu)^{b-1} d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

derive the expression for C_M.

4 Marginal and conditional distributions of joint Gaussians (50 points)

Consider a joint distribution over the variable

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y}. \end{pmatrix}$$

Given that the mean vector and covariance matrix of $\boldsymbol{z}$ are then given by:

$$\mathrm{E}[\boldsymbol{z}] = \boldsymbol{R}^{-1} \begin{pmatrix} \boldsymbol{\Lambda\mu} - \boldsymbol{A}^\top \boldsymbol{Lb} \\ \boldsymbol{Lb} \end{pmatrix}$$

and

$$\mathrm{Cov}[\boldsymbol{z}] = \boldsymbol{R}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\boldsymbol{A}^\top \\ \boldsymbol{A\Lambda}^{-1} & \boldsymbol{L}^{-1} + \boldsymbol{A\Lambda}^{-1}\boldsymbol{A}^\top \end{pmatrix}.$$

A joint Gaussian distribution given by $p(\boldsymbol{z}_a, \boldsymbol{z}_b) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$ for an arbitrary partition of the multivariate Gaussian variable $\boldsymbol{z} = (\boldsymbol{z}_a, \boldsymbol{z}_b)^\top$ where $\boldsymbol{\mu}_z = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}$ and $\boldsymbol{\Sigma}_z = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$ has a Gaussian marginal distribution $p(\boldsymbol{z}_a)$ with mean-vector $\mathrm{E}[\boldsymbol{z}_a] = \boldsymbol{\mu}_a$ and covariance matrix $\mathrm{Cov}[\boldsymbol{z}_a] = \boldsymbol{\Sigma}_{aa}$.

(a) (25 points) Use the above property to show that the marginal distribution $p(\boldsymbol{x})$ is given by $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$.

Further the Gaussian distribution $p(\boldsymbol{z}_a, \boldsymbol{z}_b)$ has a Gaussian conditional distribution $p(\boldsymbol{z}_a|\boldsymbol{z}_b)$ with the following mean vector and covariance matrix,

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{z}_b - \boldsymbol{\mu}_b)$$
$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}.$$

(b) (25 points) Use the above to show that the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ is $\mathcal{N}(\boldsymbol{y}|\boldsymbol{Ax} + \boldsymbol{b}, \boldsymbol{L}^{-1})$.