1. MLE of the Laplace Distribution

   Let X have a Laplace distribution with density

$$p(x;\mu,b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

   Suppose that n samples $x_1,\ldots,x_n$ are drawn independently according to $p(x;\mu,b)$.

   (a) Find the maximum likelihood estimate of μ.

Solution:

$$\mathcal{L}(\mu,b|D) = \prod_{i=1}^{n}\frac{1}{2b}\exp\left(-\frac{|x_i-\mu|}{b}\right) = \left(\frac{1}{2b}\right)^n \exp\left(-\frac{1}{b}\sum_{i=1}^{n}|x_i-\mu|\right)$$

$$l(\mu,b|D) = n\log\left(\frac{1}{2b}\right) - \frac{1}{b}\sum_{i=1}^{n}|x_i-\mu|$$

Taking the derivative of log-likelihood wrt μ and setting it to be equal to zero yields

$$\frac{\partial l(\mu,b|D)}{\partial \mu} = \frac{1}{b}\sum_{i=1}^{n}\operatorname{sgn}\left[x_i-\mu\right] = 0$$

The above equation is satisfied when μ_{MLE} is the sample median.

   (b) Find the maximum likelihood estimate of b.

Solution: Taking the derivative of the log likelihood, we have:

$$\frac{\partial l(\mu,b|D)}{\partial b} = -\frac{n}{b} + \frac{\sum_{i=1}^{n}|x_i-\mu|}{b^2} = 0$$

$$b_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}|x_i-\mu|$$

which is the average absolute deviation from the mean.

   (c) Assume that μ is given. Show that b_{MLE} is an unbiased estimator (to show that the estimator is unbiased, show that $\mathrm{E}\left[b_{\text{MLE}}-b\right]=0$).

Solution:

$$\mathrm{E}\left[b_{MLE}-b\right] = \mathrm{E}\left[b_{\text{MLE}}\right] - b = 0$$

To solve for $\mathrm{E}\left[b_{\text{MLE}}\right]$, we use linearity of expectation:

$$\mathrm{E}\left[b_{\text{MLE}}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[|X_i-\mu|\right] = \mathrm{E}\left[|X-\mu|\right]$$

where the last equality holds because each random variable has the same expectation. Now, define a new random variable $Z = X - \mu$. It is easy to show that Z is also Laplacian with mean $\mu_Z = 0$ and $b_Z = b$.

We now need to find $\mathrm{E}[|Z|]$. Since Laplace distribution is symmetric about the origin, we can simply double the probability density of Z and consider only the positive part to find the density of $|Z|$:

$$|Z| \sim \begin{cases} \frac{1}{b}\exp\left(-\frac{z}{b}\right) & \text{if } z \geq 0 \\ 0 & \text{o.w.} \end{cases} .$$

This is the density of an exponential random variable with parameter $\frac{1}{b}$, therefore, $E[|Z|] = b$, showing that (given μ), b_{MLE} is unbiased.

2. Transforming a Standard Normal Multivariate Gaussian

We are given a 2 dimensional multivariate Gaussian random variable Z, with mean 0 and covariance I. We want to transform this into something cooler. Find the covariance matrix of a multivariate Gaussian such that the axes x_1 and x_2 of the isocontours of the density are elliptically shaped with major/minor axis lengths in a 4:3 ratio, and the axes are rotated 45 degrees counterclockwise.

Solution:

Recall that any symmetric matrix Σ can be decomposed as $U \Lambda U^T$, where U is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of corresponding eigenvalues. Also recall that the columns of U are the directions of the ellipsoid axes and the values of $\Lambda^{\frac{1}{2}}$ correspond to the length of those axes.

(a) First, we find Λ. Recall that multiplying a random vector Z with a diagonal matrix D will scale the variances by the squares of the diagonal (the new covariance matrix of DZ is $DID^T = D^2$). The lengths of the axes of the ellipsoid are proportional to the **standard deviation** of each individual component. So, in order to scale by 3 and 4, we simply create the matrix:

$$\Lambda = \begin{pmatrix} 3^2 & 0 \\ 0 & 4^2 \end{pmatrix} = \begin{pmatrix} 9 & 0 \\ 0 & 16 \end{pmatrix}$$

Note that we are finding the spectral decomposition. In order to achieve this type of scaling we would multiply Z by $\Lambda^{\frac{1}{2}}$.

(b) Next we need to find a rotation matrix U such that it rotates the standard cartesian coordinate system 45 degrees counter clockwise. There are 2 ways to do this:

1) Remember that a rotation matrix has the form:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Plugging in $\theta = \frac{\pi}{4}$ gives us

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

2) Another way to do this is to realize that we want e_1 to be rotate 45 degrees counterclockwise. Writing that out mathematically, we have

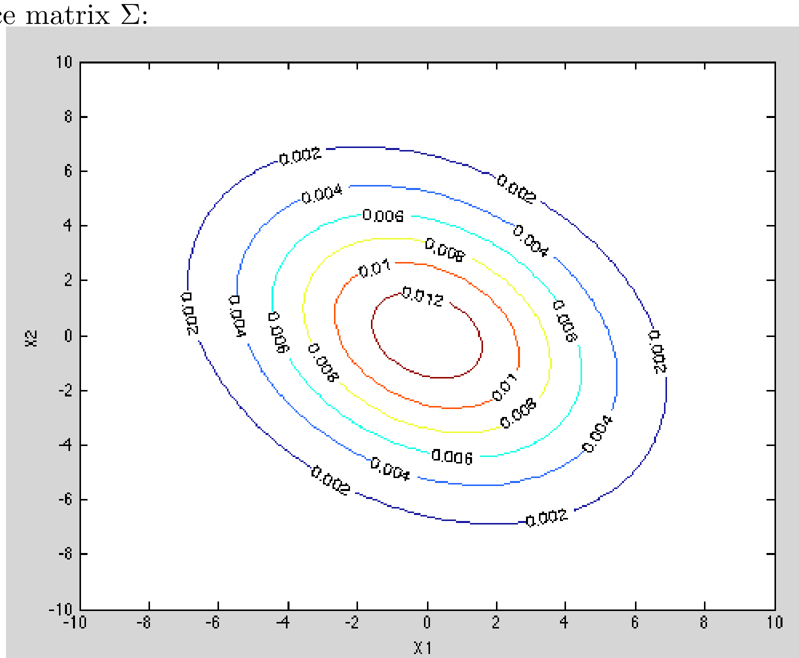
$$U * \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

where the far right hand side is the coordinates of rotating e_1 45 degrees counterclockwise on the unit circle. Doing the same for e_2 gives us the same result as the first method.

(c) Finally, we simply multiply out to find the new covariance matrix.

$$\Sigma = U \Lambda U^T = \frac{1}{2} \begin{pmatrix} 25 & -7 \\ -7 & 25 \end{pmatrix}$$

To drive this point home, here is a plot of the isocontours of the multivariate gaussian with the covariance matrix Σ:



3. Multivariate Gaussian

a) True or False

(i) If X_1 and X_2 are both normally distributed and independent, then (X_1, X_2) must have multivariate normal distribution.

Solution: True.
Since X_1 and X_2 are independent with each other, so we have

$$\begin{aligned} p(x_1, x_2) &= p(x_1)p(x_2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right) \\ &= \frac{1}{\sqrt{(2\pi)^2 \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right) \\ &= \mathcal{N}(\mu, \Sigma) \end{aligned}$$

where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

(ii) If (X_1, X_2) has multivariate normal distribution, then X_1 and X_2 are independent.

Solution: False. If the off diagonal elements of the covariance matrix Σ are not zeros, it means $\mathrm{Cov}[X_1, X_2] \neq 0$. Then they are not independent.

b) Affine transformation

$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^{\mathrm{T}}$ is a n-dimensional random vector which has multivariate normal distribution. If $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = BX + c$ is an affine transformation of X, where c is a constant $m \times 1$ vector and B is a constant $m \times n$ matrix, what is the expectation and variance of Y?

Solution:

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[BX + c] \\ &= B\mathrm{E}[X] + c \\ &= B\mu + c \\ &= \mu_Y \end{aligned}$$

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{E}\left[(Y-\mu_Y)(Y-\mu_Y)^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[(BX + c - (B\mu + c))(BX + c - (B\mu + c))^{\mathrm{T}}\right] \\ &= \mathrm{E}\left[B(X-\mu)(X-\mu)^{\mathrm{T}}B^{\mathrm{T}}\right] \\ &= B\mathrm{E}\left[(X-\mu)(X-\mu)^{\mathrm{T}}\right]B^{\mathrm{T}} \\ &= B\Sigma B^{\mathrm{T}} \end{aligned}$$

In fact, Y is normally distributed, i.e. $Y \sim \mathcal{N}\left(B\mu + c, B\Sigma B^{\mathrm{T}}\right)$, but the proof of that requires some more advanced linear algebra and probability theory.

4. [Extra for Experts] Linear Algebra

   a) Let A be a square matrix. Show that we can write A as the sum of a symmetric matrix A_+ and an antisymmetric matrix A_-:
   $$A = A_+ + A_-$$
   where $A_+ = A_+^{\mathrm{T}}$ and $A_- = -A_-^{\mathrm{T}}$.

   b) Show that if A_- is antisymmetric, then $x^{\mathrm{T}} A_- x = 0$ for all nonzero x.

   c) Show that the inverse of a positive definite matrix is positive definite.

   d) Any multivariate Gaussian distribution can be defined by two parameters, μ and Σ. It is common to assume that Σ is a positive definite matrix. Explain how we can find a Gaussian distribution corresponding to any square matrix Λ, which satisfies only $z^{\mathrm{T}}\Lambda z > 0 \; \forall z \neq 0$, but is not necessarily symmetric, and therefore not PD. Hint: make use of parts a), b), and c).

Solution:

a) Choose $A_+ = \frac{A+A^{\mathrm{T}}}{2}$ and $A_- = \frac{A-A^{\mathrm{T}}}{2}$. Notice that if A is symmetric, then $A_+ = A$ and $A_- = 0$.

b) The expression $x^T A_- x$ is a scalar, so $x^T A_- x = (x^{\mathrm{T}} A_- x)^{\mathrm{T}} = x^{\mathrm{T}} A_-^{\mathrm{T}} x = -x^{\mathrm{T}} A_- x$ (where the last step uses the fact that A_- is antisymmetric). Thus $x^{\mathrm{T}} A_- x = 0$.

c) Let B be PD. Then $(B^{-1})^{\mathrm{T}} = (B^{\mathrm{T}})^{-1} = B^{-1}$, so B^{-1} is symmetric. Further, define $y = Bz$, then $y^{\mathrm{T}} B^{-1} y = z^{\mathrm{T}} B z > 0$ for all $z \neq 0$. Since B is invertible, this also holds for all $y > 0$. Thus, B^{-1} is positive definite.

d) Let us define a distribution
$$p_\Lambda(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Lambda(x-\mu)\right) \tag{1}$$
where Z is a normalization constant and Λ is any matrix such that $z^{\mathrm{T}}\Lambda z > 0 \; \forall z \neq 0$. (Here, we make the dependence on the parameter Λ explicit while neglecting μ, since the following discussion will only focus on how the density varies as a function of Λ.)

We can see that that p_Λ depends on Λ only through the quadratic form $(x-\mu)^{\mathrm{T}}\Lambda(x-\mu)$, so by parts a) and b), we can write $(x-\mu)^{\mathrm{T}}\Lambda(x-\mu) = (x-\mu)^{\mathrm{T}}\Lambda_+(x-\mu)$, where Λ_+ is PD. By part c), we know that Λ_+^{-1} is PD, and we can write
$$p_\Lambda(x) = \mathcal{N}\left(\mu, \Lambda_+^{-1}\right).$$

In the other words, any distribution $p_\Lambda(x)$ given by (1), where Λ satisfies $z^{\mathrm{T}}\Lambda z > 0 \; \forall z \neq 0$, is a Gaussian distribution. Note that even though the antisymmetric part has no role in the quadratic part, it needs to be considered when computing the normalization constant Z.