# 1 Gaussian Isocontours

(a) Consider a linear transformation $T(x) = Ux$ where $x \in \mathbb{R}^2$ and $U \in \mathbb{R}^{2\times 2}$ that takes a vector and rotates it by 45° counterclockwise. Find the matrix $U$ that performs such a transformation. What is a special property of such a matrix? To what tranformation does $T'(x) = U^\top x$ correspond?

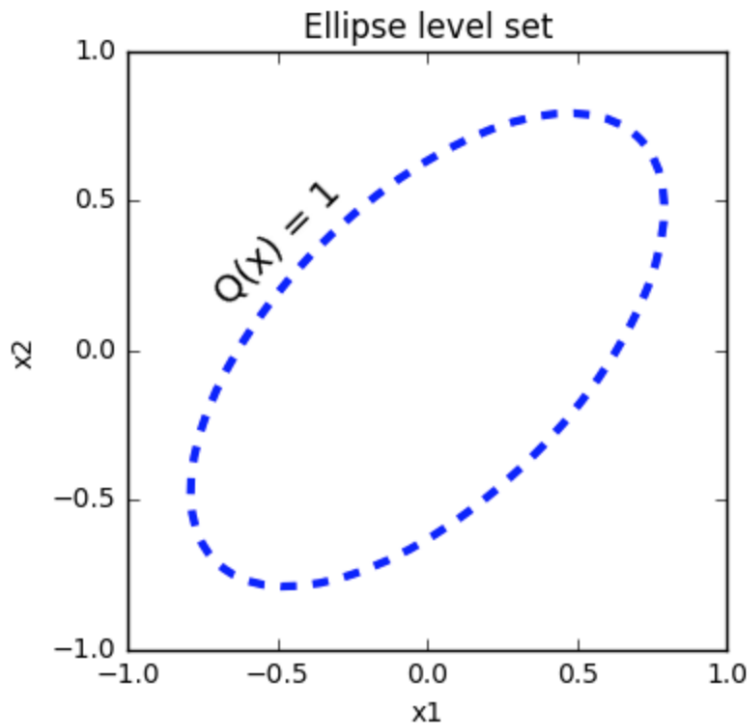
**Solution:** We want to find a matrix $U$ such that $Ue_1 = \left[\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right]^\top$ and $Ue_2 = \left[-\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right]^\top$, where $e_i$ refers to the $i$-th basis vector. Hence $U = \begin{bmatrix} \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \end{bmatrix}$. $U$ has the special property that $U^{-1} = U^\top$. Thus $T'$ corresponds to the transformation that rotates a vector by $-45°$ counterclockwise; said differently 45° clockwise.

(b) Using the matrix $U$ from the part (a), we construct a new matrix $A = U\Lambda U^\top$ where $\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. What are the eigenvalues and eigenvectors of the matrix $A$? Now consider the quadratic function $Q(x) = x^\top A^{-1} x$. Draw the level set $Q(x) = 1$.

**Solution:** Notice that since $U$ is an orthonormal matrix, $A = U\Lambda U^\top$ is the spectral decomposition of the matrix $A$. Hence, the eigenvectors and eigenvalues of $A$ are $u_1 = \left[\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right]^\top$ with corresponding eigenvalue $\lambda_1 = 2$ and $u_2 = \left[-\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right]^\top$ with corresponding eigenvalue $\lambda_2 = 1$.

Now let's draw the level set of the function $Q(x) = 1$. $Q(x) = x^\top A^{-1} x = x^\top U \Lambda^{-1} U^\top x = \frac{(x^\top u_1)^2}{\lambda_1} + \frac{(x^\top u_2)^2}{\lambda_2}$ where $u_1$ and $u_2$ are the first and second columns of the matrix $U$ respectively. Hence we have the level set $\frac{(x^\top u_1)^2}{\lambda_1} + \frac{(x^\top u_2)^2}{\lambda_2} = 1$ which is an ellipse centered at the origin with the major axis given by $\sqrt{\lambda_1} u_1$ and the minor axis given by $\sqrt{\lambda_2} u_2$.

DIS5, ©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission. 




(c) Using the result from part (b) show that the isocontours of a multivariate Gaussian $X \sim N(\mu, \Sigma)$ where $\Sigma \succ 0$ are also ellipses.

*Hint:* Recall that the density of a multivariate Gaussian is given by

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

**Solution:** Notice that the multivariate Gaussian density has a quadratic term in the exponent. The isocontours for our density are given by $\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) = c$.

$$\begin{aligned}
\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) &= c \\
\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu) &= \log\left(\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\right) - \log(c) \\
(x-\mu)^\top \Sigma^{-1}(x-\mu) &= 2(\log\left(\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\right) - \log(c))
\end{aligned}$$

Now observe that the RHS of the equation is simply a constant and hence the isocontours will look like ellipses with major axes in the direction of the largest eigenvalue $\Sigma$ and a minor axis in the direction of the smallest eigenvalue of $\Sigma$.

For the remainder of this problem, we will explore the shape of quadratic forms by examining the eigen-structure of the Hessian matrix. Recall that the Hessian $H \in \mathbb{R}^{d \times d}$ of a function

DIS5, ©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission. 

$f : \mathbb{R}^d \to \mathbb{R}$ is the matrix of second derivatives $H_{i,j} = \frac{\partial f}{\partial x_i x_j}$ of the function. The eigen-structure of $H$ contains information about the curvature of $f$.

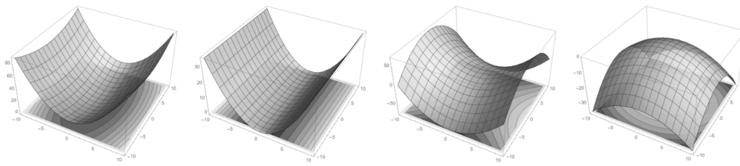
(d) Suppose you are given the a quadratic function $Q(x) = \frac{1}{2} x^\top A x$ where $x \in \mathbb{R}^2$ and $A \in \mathbb{R}^{2\times 2}$ is a symmetric matrix. What is the Hessian of $Q$?

**Solution:**

$$\nabla_x f(x) = \frac{1}{2}(A + A^\top)x = Ax$$
$$\nabla_x^2 f(x) = \nabla_x Ax = A^\top = A$$

©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission.

(e) We will now think about how the eigen-structure of the Hessian matrix affects the shape of the $Q(x)$. Recall that by the Spectral Theorem, $A$ has two real eigenvalues. Match each of the following cases, to the appropriate plot of $Q(x)$. How does the magnitude of the eigenvectors affect your sketch?



(a) $\lambda_1(A), \lambda_2(A) > 0$

(b) $\lambda_1(A) > 0, \lambda_2(A) = 0$

(c) $\lambda_1(A) > 0, \lambda_2(A) < 0$

(d) $\lambda_1(A), \lambda_2(A) < 0$

**Solution:**

(a) Since both eigenvalues are strictly positive, $A \succ 0$ which implies that $x^\top A x > 0, \forall x \neq \mathbf{0}$. As a result, we can see that the function has a unique global minimum at $x = \mathbf{0}$. Let's examine the curvature around the origin. Our goal is to look at all possible directions and understand what the curvature looks like for each one. Recall from HW2 that $\max_{\|u\|=1} u^\top A u = \lambda_{\max}(A)$ and $\min_{\|u\|=1} u^\top A u = \lambda_{\min}(A)$. As a result, if we move just a little bit away from 0 in the direction of the eigenvector corresponding to $\lambda_{\max}(A)$, we will experience the steepest curvature. On the other hand, if we move in the direction of the eigenvector corresponding to $\lambda_{\min}(A)$ we will experience the least curvature. The resulting quadratic should look something like the surface in the first figure from left.

(b) Figure 2

(c) Figure 3

(d) Figure 4

DIS5, ©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission. 

## 2 Linear Discriminant Analysis

In this question, we will explore some of the mechanics of LDA and understand why it produces a linear decision boundary in the case where the covariance matrix is anisotropic.

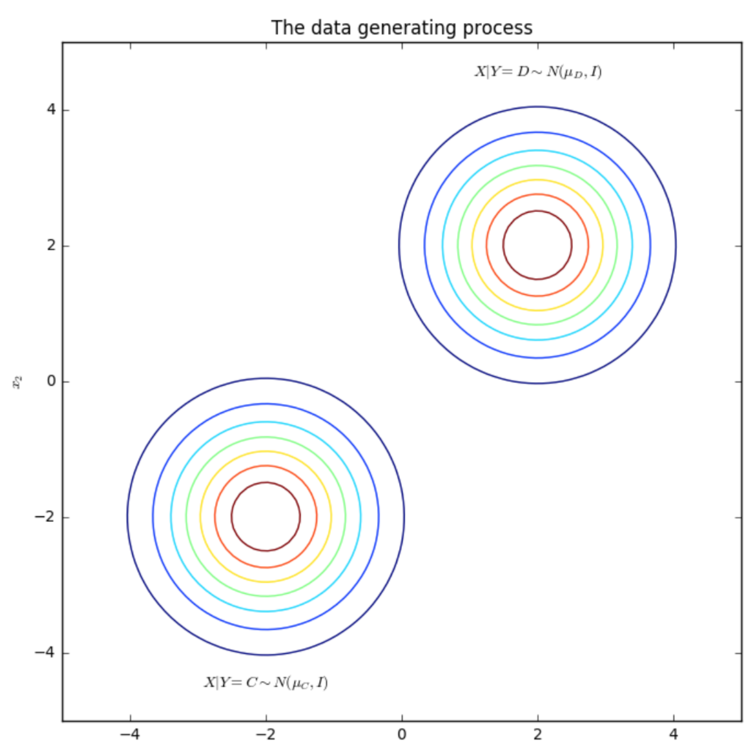
(a) Suppose $\Sigma = \mathrm{Cov}(X)$ is the covariance matrix of random vector $X \in \mathbb{R}^d$. Prove that $\mathrm{Cov}(AX) = A\Sigma A^\top$.

**Solution:**

$$\begin{aligned}\mathrm{Cov}(AX) &= \mathrm{E}[(AX - A\,\mathrm{E}[X])(AX - A\,\mathrm{E}[X])^\top] = \mathrm{E}[A(X - \mathrm{E}[X])(X - \mathrm{E}[X])^\top A^\top]] \\ &= A\,\mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])^\top]A^\top \\ &= A\Sigma A^\top.\end{aligned}$$

(b) Suppose you have a binary classification problem. You are given a design matrix $X \in \mathbb{R}^{n\times 2}$ and a set of labels $y \in \mathbb{R}^n$ such that $y_i \in \{C, D\}$. A genie comes to you and gives you the following additional information about the process that generated the data.

- The two classes have identical priors $P(Y = C) = P(Y = D) = \frac{1}{2}$
- The class conditional-densities are $X|Y = C \sim N(\mu_C, I)$ and $X|Y = D \sim N(\mu_D, I)$ where $\mu_C = \begin{bmatrix}-2\\-2\end{bmatrix}$, $\mu_D = \begin{bmatrix}2\\2\end{bmatrix}$.




DIS5, ©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission. 

We can recognize this problem as a special case of LDA where the two classes have an equal prior probability and the common covariance matrix is simply the identity. Use Bayes' Decision Rule to construct a decision boundary for this problem.

*Hint: You may want to start by drawing the decision boundary on the plot provided. Does the result line up with your intuition?*

**Solution:** The decision boundary simply ends up being the perpendicular bisector of the line connecting $\mu_0$ and $\mu_1$. To find the decision boundary consider:

$$\begin{aligned}
P(Y = D|X) &= P(Y = C|X) \\
P(X|Y = D)P(Y = D) &= P(X|Y = C)P(Y = C) \\
P(X|Y = D) &= P(X|Y = C) \\
\frac{1}{\sqrt{2\pi|I|}} \exp\left(-\frac{1}{2}(x-\mu_C)^\top I(x-\mu_C)\right) &= \frac{1}{\sqrt{2\pi|I|}} \exp\left(-\frac{1}{2}(x-\mu_D)^\top I(x-\mu_D)\right) \\
\|x-\mu_C\|_2^2 &= \|x-\mu_D\|_2^2
\end{aligned}$$

Hence the decision boundary contains all the points that are equidistant from the two class means. The set of points $h = \{x \in \mathbb{R}^2 : \|x-\mu_C\|_2^2 = \|x-\mu_D\|_2^2\}$ is exactly the perpendicular bisector of the line that connects the two means in the picture. That is, $h$ is the plane that is orthogonal to the vector $\mu_D - \mu_C$ and passes through the point $\frac{\mu_C+\mu_D}{2}$.

(c) Now we will try to use this intuition to explain why the decision boundary also has to be linear when the class-conditinal densities have a more general covariance matrix $\Sigma \geq 0$.

Assume that we are given the same setup as in the previous part, but this time the covariance matrix is some known $\Sigma \geq 0$ instead of the identity matrix. Find a linear transformation such that the class-conditional distributions are isotropic Gaussians in the transformed space. What is the decision boundary in the transformed space? What does that boundary correspond to in the original space?

*Hint: The result you proved in Problem 1 may be useful.*

**Solution:** $T(x) = \Sigma^{-\frac{1}{2}}x$ and $\bar{x} = T(x)$ for all $x$ in the original space.

The decision boundary has to be some plane in the transformed space of the form $w^\top \bar{x}$ (the decision boundary goes through the origin because of the specific values $\mu_C$ and $\mu_D$ that we were given). In that case, the boundary in the original space will be $\bar{w}^\top x$ for $\bar{w} = \Sigma^{\frac{1}{2}}w$ which is itself an affine decision function.

 ©UCB CS 189, Spring 2019. All Rights Reserved. This may not be publicly shared without explicit permission. 