CS 189 Introduction to Machine Learning
Spring 2019 Jonathan Shewchuk

DIS5

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?

(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 $Q(x) = 1$.

(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.

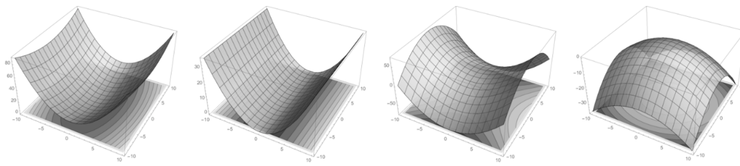
Hint: Recall that the density of a multivariate Gaussian is given by

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

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 $f : \mathbb{R}^d \to \mathbb{R}$ 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?

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

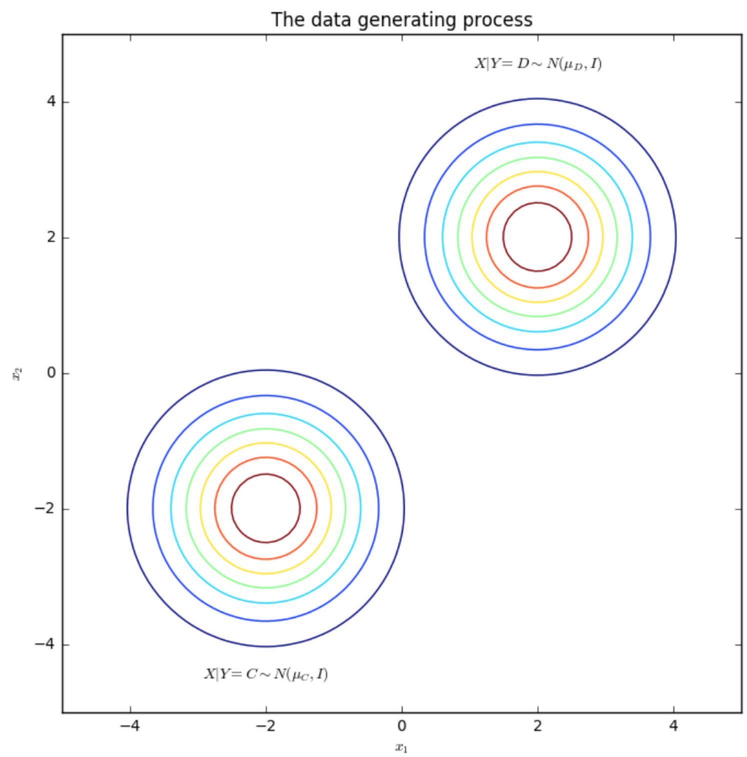
2 Linear Discriminant Analysis

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 $\mathrm{Cov}(AX) = A\Sigma A^\top$.

(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 $\mu_C = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$, $\mu_D = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.



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?

(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 $\Sigma \succeq 0$ 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.