CS 189 Spring 2023 | Introduction to Machine Learning 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. What does $T'(x) = U^\top x$ mean?

(b) Now let's verify some basic properties of this matrix U. What is the determinant of U? Does U^{-1} exist? If it does, can you express U^{-1} in terms of U? Usually, we refer to matrices that have the above properties as the Special Orthogonal group, or SO(2), as the matrix is 2 by 2. They are also called rotational matrices.

(c) With the matrix U from 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 $Q(x) = 1$.

(d) Using the result from part (c), show that the isocontours of a multivariate Gaussian $X \sim N(\mu, \Sigma)$ where $\Sigma \succ 0$ are also ellipses. Isocontour means a set of X that has the same probability of being sampled. For now, you can assume that you can read off probabilities from continuous probability density functions.

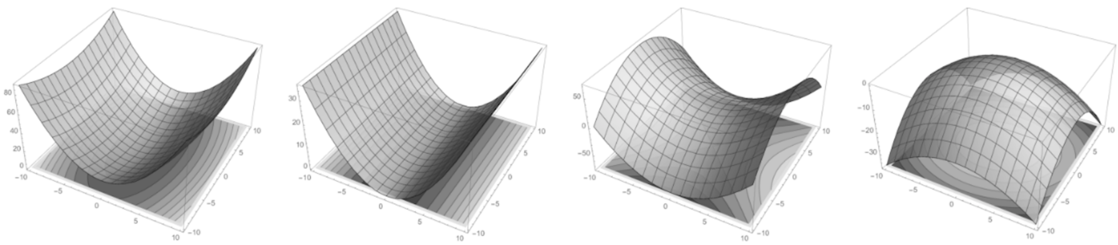
Hint: Recall that the density of a multivariate Gaussian is

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

For the remainder of this problem, we will explore the shape of quadratic forms by examining the eigenstructure 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 eigenstructure of H contains information about the curvature of f.

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

(f) We will now think about how the eigenstructure of the Hessian matrix affects the shape of $Q(x)$. 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 eigenvalues affect your sketch? For each of these 4 cases, is there an unique local optimum?



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

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

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

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

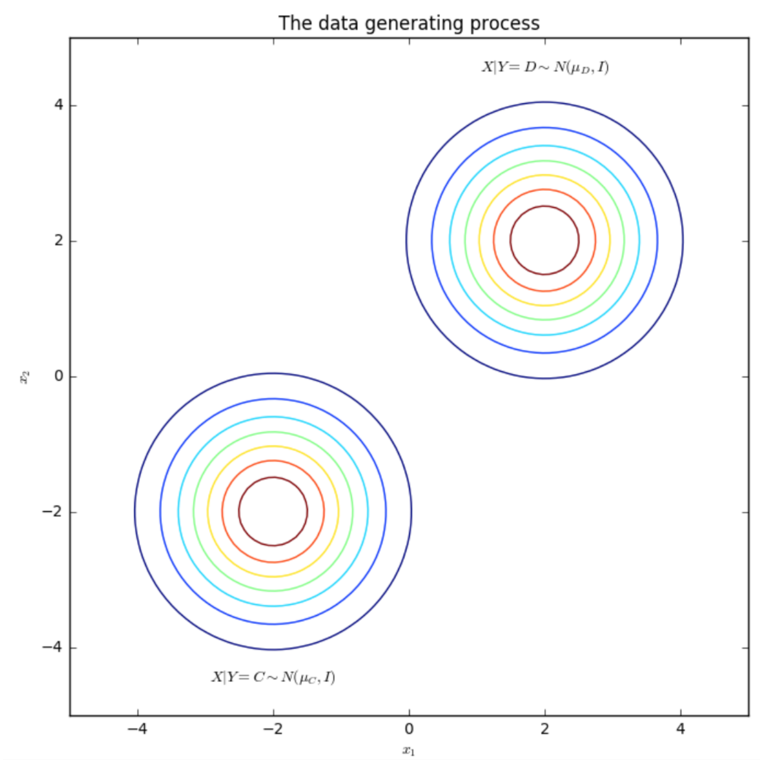
2 Linear Discriminant Analysis

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

(b) Suppose you have a binary classification problem with $x \in \mathbb{R}^2$, and you already know the data generating process.

- 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 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. Recall the optimal Bayes optimal decision boundary from the last discussion worksheet, which is the perpendicular bisector of the line connecting μ_0 and μ_1.

Now we will use this intuition to explain why the decision boundary also has to be linear when the class-conditional densities have a more general covariance matrix $\Sigma \succeq 0$.

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.

(c) Consider the case below where $(X|Y = C) \sim N(\mu_C, \alpha I)$. Characterize alpha and roughly indicate the shape of the decision boundary.