CS 189 Introduction to Machine Learning
Spring 2021 Jonathan Shewchuk **Exam Prep 2 Solutions**

This exam-prep discussion section covers anisotropic normal distributions, Quadratic and Linear Discriminant Analyses, and some of their extensions.

1 Multiple Choice

(m) [3 pts] In LDA/QDA, what are the effects of modifying the sample covariance matrix as $\tilde{\Sigma} = (1 - \lambda)\Sigma + \lambda I$, where $0 < \lambda < 1$?

- ● $\tilde{\Sigma}$ is positive definite
- ● $\tilde{\Sigma}$ is invertible
- ○ Increases the eigenvalues of Σ by λ
- ● The isocontours of the quadratic form of $\tilde{\Sigma}$ are closer to spherical

(h) [3 pts] We are using **linear discriminant analysis** to classify points $x \in \mathbb{R}^d$ into **three** different classes. Let S be the set of points in $\mathbb{R}^d$ that our trained model classifies as belonging to the first class. Which of the following are true?

- ○ The decision boundary of S is always a hyperplane
- ● S can be the whole space $\mathbb{R}^d$
- ● The decision boundary of S is always a subset of a union of hyperplanes
- ● S is always connected (that is, every pair of points in S is connected by a path in S)

Top left: Given that we have three classes, S is defined by two linear inequalities, and therefore its boundary may not be a hyperplane.

Bottom left: Given that S is defined as the points satisfying a set of inequalities, its boundary is a subset of the hyperplanes defined by each of the linear inequalities.

Top right: If the prior for the first class is high enough, the probability of that class could be higher everywhere, and hence S would be the whole space. For example, take $\mu_1 = \mu_2 = \mu_3$ and $\pi_1 > \pi_2 = \pi_3$.

Bottom right: S is a convex polytope defined by the intersection of half-spaces (i.e. the points satisfying a set of linear inequalities). This is a convex set, and therefore it is connected.

(o) [3 pts] Suppose you have a **multivariate normal distribution** with a positive definite covariance matrix Σ. Consider a second multivariate Gaussian distribution whose covariance matrix is $\kappa\Sigma$, where $\kappa = \cos\theta > 0$. Which of the following statements are true about the ellipsoidal isocontours of the second distribution, compared to the first distribution?

- ○ The principal axes of the ellipsoids would be rotated by θ
- ○ The principal axes (radii) of the ellipsoids will be scaled by $1/\kappa$
- ○ The principal axes (radii) of the ellipsoids will be scaled by κ
- ● The principal axes (radii) of the ellipsoids will be scaled by $\sqrt{\kappa}$

Multiplying Σ by κ multiplies the eigenvalues by κ. The axes of the ellipsoids are scaled by the square roots of the eigenvalues of Σ.

(s) [3 pts] Suppose you have a sample in which each point has d features and comes from class C or class D. The class conditional distributions are $(X_i|y_i = C) \sim N(\mu_C, \sigma_C^2)$ and $(X_i|y_i = D) \sim N(\mu_D, \sigma_D^2)$ for unknown values $\mu_C, \mu_D \in \mathbb{R}^d$ and $\sigma_C^2, \sigma_D^2 \in \mathbb{R}$. The class priors are π_C and π_D. We use 0-1 loss.

● If $\pi_C = \pi_D$ and $\sigma_C = \sigma_D$, then the Bayes decision rule assigns a test point z to the class whose mean is closest to z.

● If $\pi_C = \pi_D$, then the Bayes decision rule is $r^*(z) = \text{argmin}_{A\in\{C,D\}} \left(|z - \mu_A|^2/(2\sigma_A^2) + d \ln \sigma_A\right)$

● If $\sigma_C = \sigma_D$, then the Bayes decision boundary is always linear.

○ If $\sigma_C = \sigma_D$, then QDA will always produce a linear decision boundary when you fit it to your sample.

2 Quadratics and Gaussian Isocontours (Spring 2016)

(a) [4 pts] Write the 2×2 matrix Σ whose unit eigenvectors are $\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ with eigenvalue 1 and $\begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$ with eigenvalue 4. Write out **both** the eigendecomposition of Σ and the final 2×2 matrix Σ.

$$\Sigma = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 17/5 & -6/5 \\ -6/5 & 8/5 \end{bmatrix}.$$

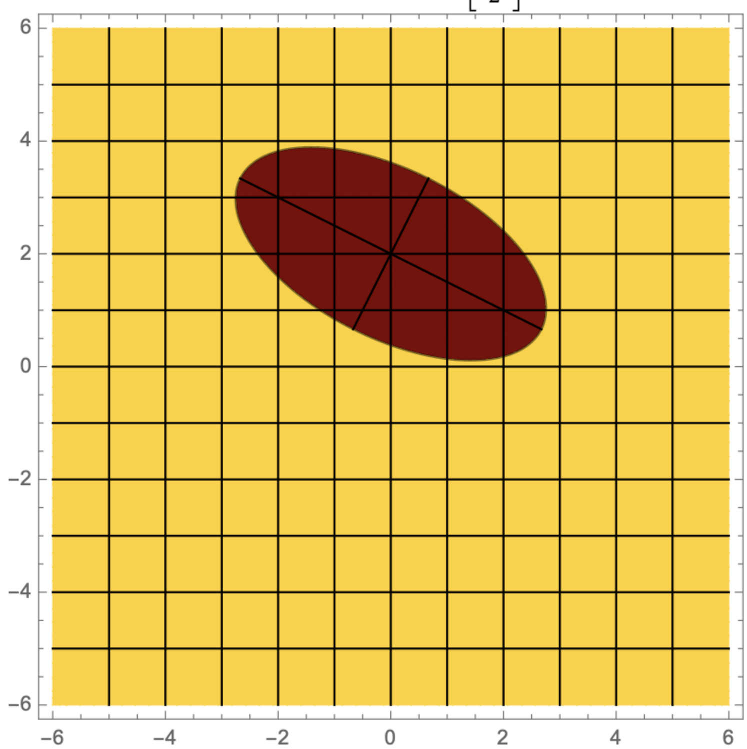
(b) [3 pts] Write the symmetric square root $\Sigma^{1/2}$ of Σ. (The eigendecomposition is optional, but it might earn you partial credit if you get $\Sigma^{1/2}$ wrong.)

$$\Sigma^{1/2} = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 9/5 & -2/5 \\ -2/5 & 6/5 \end{bmatrix}.$$

(c) [3 pts] Consider the bivariate Gaussian distribution $X \sim \mathcal{N}(\mu, \Sigma)$. Let $P(X = \mathbf{x})$ be its probability distribution function (PDF). Write the formula for the isocontour $P(\mathbf{x}) = e^{-\sqrt{5}/2}/(4\pi)$, substitute in the value of the determinant $|\Sigma|$ from part (a) (but leave μ and Σ^{-1} as variables), and simplify the formula as much as you can.

$$\begin{aligned} \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{(\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu)}{2}\right) &= \frac{e^{-\sqrt{5}/2}}{4\pi} \\ (\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu) &= \sqrt{5} \end{aligned}$$

(d) [5 pts] Draw the isocontour $P(\mathbf{x}) = e^{-\sqrt{5}/2}/(4\pi)$ where $\mu = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ and Σ is given in part (a).



3 Discriminant Analysis (Spring 2016)

Let's derive the decision boundary when one class is Gaussian and the other class is exponential. Our feature space is one-dimensional ($d = 1$), so the decision boundary is a small set of points.

We have two classes, named N for normal and E for exponential. For the former class ($Y = N$), the prior probability is $\pi_N = P(Y = N) = \frac{\sqrt{2\pi}}{1+\sqrt{2\pi}}$ and the class conditional $P(X|Y = N)$ has the normal distribution $\mathcal{N}(0, \sigma^2)$. For the latter, the prior probability is $\pi_E = P(Y = E) = \frac{1}{1+\sqrt{2\pi}}$ and the class conditional has the exponential distribution

$$P(X = x|Y = E) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Write an equation in x for the decision boundary. (Only the positive solutions of your equation will be relevant; ignore all $x < 0$.) Use the 0-1 loss function. Simplify the equation until it is quadratic in x. (You don't need to solve the quadratic equation. It should contain the constants σ and λ. Ignore the fact that 0 might or might not also be a point in the decision boundary.) **Show your work**, starting from the posterior probabilities.

Ignoring the possibility of $x = 0$, the decision boundary is the set of positive solutions to

$$\begin{aligned} P(Y = N|X = x) &= P(Y = E|X = x) \\ \frac{P(X = x|Y = N)P(Y = N)}{P(X = x)} &= \frac{P(X = x|Y = E)P(Y = E)}{P(X = x)} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{\sqrt{2\pi}}{1+\sqrt{2\pi}} &= \lambda e^{-\lambda x} \frac{1}{1+\sqrt{2\pi}} \\ -\ln\sigma - \frac{x^2}{2\sigma^2} &= \ln\lambda - \lambda x \\ 0 &= \frac{x^2}{2\sigma^2} - \lambda x + \ln\lambda + \ln\sigma. \end{aligned}$$

Note that the last term can be abbreviated to $\ln(\lambda\sigma)$. The last line above is not necessary for full credit; the second-last line counts as a "quadratic equation." The first line of math also is not necessary for full credit, but Bayes' Theorem must implicitly be present.