Homework Five, due Thu 3/9 CSE 250B

- Hand in your homework in hard copy.
- Present your results clearly and succinctly using tables and graphs as appropriate. Discuss your results in precise and lucid prose. Content is king, but looks matter too!
- Please turn in code *only* for the functions specifically mentioned in the exercise (plus any helper functions which they call). The code for these functions should be emailed to `<seano@cs.ucsd.edu>`. Please include CSE 250B HW 3 in the subject line of your email.

1. **Generative models example.** Your two-dimensional data set consists of two classes which are present in equal proportion and have the following first- and second-order statistics.

   - Class 1: mean $\begin{pmatrix}4\\8\end{pmatrix}$, covariance $\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}$
   - Class 2: mean $\begin{pmatrix}8\\4\end{pmatrix}$, covariance $\begin{pmatrix}2 & 0\\0 & 1\end{pmatrix}$

   (a) Suppose you fit a Gaussian to each class. Sketch a couple of contour lines for each cluster.

   (b) Derive the formula for the Bayes-optimal decision boundary and plot it.

2. **Cost-sensitive classification with generative models.** You are designing a pregnancy test based on a new set of features. You have a generative model with the following profile:

   - Class 0 (not pregnant): density $P_0(x)$, mixing weight π_0
   - Class 1 (pregnant): density $P_1(x)$, mixing weight π_1

   Based on this model, you want to decide whether a person with features x is pregnant or not. However, in this case the cost of misclassification is not symmetric: here's the cost you incur if you predict wrong.

COST MATRIX		True label 0	True label 1
Prediction	0	0	100
	1	10	0

   Assuming your model is perfectly correct, how should you predict on x, to minimize *expected cost*? (Give a prediction rule based on $P_0(x)$, $P_1(x)$, π_0, π_1.)

3. **Linear regression.** In class we have so far been focusing exclusively on *classification* problems, in which the variable y to be predicted takes on a discrete, finite set of values and there isn't necessarily any ordering between these values. In a *regression* problem, y is ordered and typically continuous.

   Suppose you have a data set $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ where the $x_i \in \mathbf{R}^d$ and the $y_i \in \mathbf{R}$. You want to find a linear function $w \in \mathbf{R}^d$ so that $y_i \approx w \cdot x_i$. Specifically, you want to pick the w which minimizes the least-squares cost:

   $$L(w) \;=\; \sum_{i=1}^{m} (y_i - w \cdot x_i)^2.$$

   (a) What is the first derivative ∇L and Hessian H at a point w?

   (b) Show that $L(w)$ is convex in w.

   (c) What is the gradient descent update rule?

   (d) What is the Newton-Raphson update rule?