DATA 1010
Problem Set 9
Due 09 November 2018 at 11 PM

Problem 1

The *Epanechnikov* kernel is defined by

$$D(u) = \frac{3}{4}(1-u^2)\mathbf{1}_{|u|\leq 1}.$$

(a) Is D continuous? Is it differentiable? Is it twice differentiable?

(b) Is the tri-cube weight function continuous? Is it differentiable? Is it twice differentiable?

Feel free to use technology to perform the symbolic differentiation in this problem.

Problem 2

Consider two random variables X and Y whose joint distribution has probability mass of $\frac{1}{n}$ at each of the n points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ in $\mathbb{R}^2$. Show that the covariance matrix of X and Y is equal to

$$\frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} x_i - \overline{x} \\ y_i - \overline{y} \end{bmatrix} \begin{bmatrix} x_i - \overline{x} & y_i - \overline{y} \end{bmatrix}.$$

where $\overline{x} = (x_1 + \cdots + x_n)/n$ and $\overline{y} = (y_1 + \cdots + y_n)/n$.

Problem 3

Suppose that the distribution of (X, Y) is uniform on the union of the rectangles $[0,3] \times [0,1]$ and $[0,3] \times [2,3]$.

(a) Find the regression function $r(x) = \mathbb{E}[Y \mid X = x]$.

(b) Generate 1000 samples from this distribution.

(c) Using the samples you generated, estimate the regression function $r(x)$ using a kernel density estimator with bandwidth λ selected by cross-validation.

(d) Find the Nadaraya-Watson estimator of $r(x)$, with λ selected by RSS cross-validation.

Problem 4

The value of the Nadaraya-Watson estimator $\widehat{r}(x)$ can be thought of as the constant function which minimizes the weighted residual sum of squares, with the weight applied to each sample according to its horizontal distance from x. This optimization problem must be solved on a per-x basis, since the weights change for different values of x.

(a) Using the exam scores example, adjust this procedure by fitting a *line* for each point. In other words, for each $x \in [0, 20]$, find β_0 and β_1 such that

$$\sum_{i=1}^{n} D_\lambda(x - x_i)(y_i - \beta_0 - \beta_1 x_i)^2$$

is minimized (note that you are finding a new β_0 and β_1 for each value of x). Then set $\widehat{r}(x) = \beta_0 + \beta_1 x$. You may do this optimization using the `Optim` package.

(b) Plot the new estimator. Does it curl up at the ends of the interval like the Nadaraya-Watson estimator? Explain your intuition for why this is the case.

Problem 5

(a) Find the variance of the uniform distribution on the interval $[0, 10]$.

(b) Generate 10 independent samples from the uniform distribution, calculate the average $\overline{X}$ for those samples, and estimate the variance as $\widehat{V} = \frac{1}{n}\sum_{i=1}^{10}(X_i - \overline{X})^2$. Package this whole process as a function, and call it a million times to find the mean of $\widehat{V}$.

(c) Which is larger, the answer to (a) or the answer to (b)? Calculate the percent error.

Problem 6

In this problem, we will implement a classifier for the flower data based on kernel density estimation.

(a) For each color, find the cross-validation kernel density estimator for the set of flowers of that color.

(b) Substitute your estimates into

$$m_{(X_1,X_2)=(x_1,x_2)}(c) = \frac{p_c f_c(x_1, x_2)}{\sum_{d\in\{\mathrm{R,G,B}\}} p_d f_d(x_1, x_2)}, \tag{1.1.4}$$

to obtain a classifier (in the form of a Julia function).

(c) Make a plot similar to Figure 1.13 for this classifier.

Problem 7

(a) Show that if f_1, f_2 are different multivariate normal densities on $\mathbb{R}^2$, then the set of points (x, y) for which $f_1(x, y) = f_2(x, y)$ is a line or a conic section (in other words, it is the solution set of a linear or quadratic equation).

(b) Show that if the covariance matrices for the two densities are the same, then the solution set of $f_1(x, y) = f_2(x, y)$ is a line.

You might find this code block helpful.

```
using SymPy
@vars x y μ₁ μ₂ a b c real=true
Σ⁻¹ = [a c; c b]
v = [x - μ₁, y - μ₂]
expand(v' * Σ⁻¹ * v)
```

Problem 8

Consider a binary classification problem where conditional density of class 0 is uniform on the left half of the unit square and the conditional density of class 1 is uniform on the right half of the unit square. Devise a learner which is **maximally overfit** in the sense that its training error is zero and its generalization error is maximal (that is, the learner gets every classification wrong, with probability 1).