CS 5526 – Virginia Tech
Homework 3

Brennon Bortz
(PID: brennon, Campus: Blacksburg)

15 April 2015

Code for this assignment is available at `https://github.com/brennon/cs5526-hw3`.

Written Problems

1. (20 points) Consider the unit hypersphere (with radius $r = 1$). Inside the hypersphere inscribe a hypercube (i.e., the largest hypercube you can fit inside the hypersphere). An example in two dimensions is shown in Figure 6.12 (in the text). Answer the following questions:

   (a) Derive an expression for the volume of the inscribed hypercube for any given dimensionality d. Derive the expression for one, two, and three dimensions, and then generalize to higher dimensions.

   Solution: For a hypersphere with radius $r = 1$, r will be half the diagonal of the largest hypercube that can be inscribed within this hypersphere. Thus, the length of the edge of the largest hypercube that can be inscribed within such a hypersphere is $2 \cdot \frac{1}{\sqrt{2}} = \frac{2}{\sqrt{2}}$. The volume of a hypercube with edge length l is given as

   $$\mathrm{vol}(H_d(l)) = l^d \tag{1}$$

   Thus, the volume of the largest hypercube that can be inscribed within a hypersphere with radius $r = 1$ in one dimension is

   $$\mathrm{vol}\left(H_1\left(\frac{2}{\sqrt{2}}\right)\right) = \left(\frac{2}{\sqrt{2}}\right)^1 = \frac{2}{\sqrt{2}} \tag{2}$$

   For two and three dimensions, we have

   $$\mathrm{vol}\left(H_2\left(\frac{2}{\sqrt{2}}\right)\right) = \left(\frac{2}{\sqrt{2}}\right)^2 = \frac{4}{2} = 2 \tag{3}$$

   and

   $$\mathrm{vol}\left(H_3\left(\frac{2}{\sqrt{2}}\right)\right) = \left(\frac{2}{\sqrt{2}}\right)^3 = \frac{8}{2\sqrt{2}} = \frac{4}{\sqrt{2}} \tag{4}$$

In general, for d dimensions, we have

$$\text{vol}\left(H_d\left(\frac{2}{\sqrt{2}}\right)\right) = \left(\frac{2}{\sqrt{2}}\right)^d \tag{5}$$

(b) What happens to the ratio of the volume of the inscribed hypercube to the volume of the enclosing hyperspehere as $d \to \infty$? Again, give the ratio in one, two and three dimensions, and then generalize.

Solution: The volume of a hypersphere in d dimensions with radius r, as given by Equation (6.4) in the text as

$$\text{vol}(S_d(r)) = K_d r^d = \left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d \tag{6}$$

As such, the ratio of the volume of the largest hypercube that can be inscribed within a hypersphere with radius $r = 1$ in one dimension is

$$\frac{\text{vol}(H_1(\frac{2}{\sqrt{2}})}{\text{vol}(S_1(1))} = \frac{\left(\frac{2}{\sqrt{2}}\right)^1}{\left(\frac{\pi^{1/2}}{\Gamma(1/2+1)}\right)1^1} = \frac{\frac{2}{\sqrt{2}}}{\frac{\sqrt{\pi}}{\frac{\sqrt{\pi}}{2}}} = \frac{\frac{2}{\sqrt{2}}}{\frac{1}{\frac{1}{2}}} = \frac{\frac{2}{\sqrt{2}}}{2} = \frac{1}{\sqrt{2}} \tag{7}$$

In two and three dimensions, we have

$$\frac{\text{vol}(H_2(\frac{2}{\sqrt{2}})}{\text{vol}(S_2(1))} = \frac{2}{\left(\frac{\pi^{\frac{2}{2}}}{\Gamma(\frac{2}{2}+1)}\right)1^2} = \frac{2}{\frac{\pi}{\Gamma(2)}} = \frac{2}{\frac{\pi}{1}} = \frac{2}{\pi} \tag{8}$$

and

$$\frac{\text{vol}(H_3(\frac{2}{\sqrt{2}})}{\text{vol}(S_3(1))} = \frac{\frac{4}{\sqrt{2}}}{\left(\frac{\pi^{\frac{3}{2}}}{\Gamma(\frac{3}{2}+1)}\right)1^3} = \frac{\frac{4}{\sqrt{2}}}{\frac{\pi^{\frac{3}{2}}}{\frac{3}{4}\sqrt{\pi}}} = \frac{\frac{4}{\sqrt{2}}}{\frac{4}{3}\pi} = \frac{4}{\sqrt{2}} \cdot \frac{3}{4\pi} = \frac{12}{4\pi\sqrt{2}} = \frac{3}{\pi\sqrt{2}} \tag{9}$$

In general, we have

$$\frac{\text{vol}(H_d(2/\sqrt{2}))}{\text{vol}(S_d(1))} = \frac{(2/\sqrt{2})^d}{\left(\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}\right)1^d} \tag{10}$$

As $d \to \infty$, the numerator also approaches inf. The denominator, on the other hand, grows smaller and smaller, as Γ is a factorial function. Thus, as $d \to \infty$, the ratio itself also approaches ∞.

2. (20 points) Consider the data in Table 7.1 (in text). Define the kernel function as follows $K(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2$. Answer the following questions:

(a) Compute the kernel matrix $\mathbf{K}$.
Solution:

$$\begin{aligned} K(\mathbf{x}_4, \mathbf{x}_1) &= K(\mathbf{x}_1, \mathbf{x}_4) = \|\begin{pmatrix} 4 & 2.9 \end{pmatrix} - \begin{pmatrix} 2.5 & 1 \end{pmatrix}\|^2 \\ &= \|\begin{pmatrix} 1.5 & 1.9 \end{pmatrix}\|^2 \\ &= \begin{pmatrix} 1.5 & 1.9 \end{pmatrix} \begin{pmatrix} 1.5 \\ 1.9 \end{pmatrix} \\ &= 5.86 \end{aligned}$$

$$\begin{aligned} K(\mathbf{x}_7, \mathbf{x}_1) &= K(\mathbf{x}_1, \mathbf{x}_7) = \|\begin{pmatrix} 4 & 2.9 \end{pmatrix} - \begin{pmatrix} 3.5 & 4 \end{pmatrix}\|^2 \\ &= \|\begin{pmatrix} 0.5 & -1.1 \end{pmatrix}\|^2 \\ &= \begin{pmatrix} 0.5 & -1.1 \end{pmatrix} \begin{pmatrix} 0.5 \\ -1.1 \end{pmatrix} \\ &= 1.46 \end{aligned}$$

$$\begin{aligned} K(\mathbf{x}_9, \mathbf{x}_1) &= K(\mathbf{x}_1, \mathbf{x}_9) = \|\begin{pmatrix} 4 & 2.9 \end{pmatrix} - \begin{pmatrix} 2 & 2.1 \end{pmatrix}\|^2 \\ &= \|\begin{pmatrix} 2 & 0.8 \end{pmatrix}\|^2 \\ &= \begin{pmatrix} 2 & 0.8 \end{pmatrix} \begin{pmatrix} 2 \\ 0.8 \end{pmatrix} \\ &= 4.64 \end{aligned}$$

$$\begin{aligned} K(\mathbf{x}_7, \mathbf{x}_4) &= K(\mathbf{x}_4, \mathbf{x}_7) = \|\begin{pmatrix} 2.5 & 1 \end{pmatrix} - \begin{pmatrix} 3.5 & 4 \end{pmatrix}\|^2 \\ &= \|\begin{pmatrix} -1 & -3 \end{pmatrix}\|^2 \\ &= \begin{pmatrix} -1 & -3 \end{pmatrix} \begin{pmatrix} -1 \\ -3 \end{pmatrix} \\ &= 10 \end{aligned}$$

$$\begin{aligned} K(\mathbf{x}_9, \mathbf{x}_4) &= K(\mathbf{x}_4, \mathbf{x}_9) = \|\begin{pmatrix} 2.5 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 2.1 \end{pmatrix}\|^2 \\ &= \|\begin{pmatrix} 0.5 & -1.1 \end{pmatrix}\|^2 \\ &= \begin{pmatrix} 0.5 & -1.1 \end{pmatrix} \begin{pmatrix} 0.5 \\ -1.1 \end{pmatrix} \\ &= 1.46 \end{aligned}$$

$$
\begin{aligned}
K(\mathbf{x}_9, \mathbf{x}_7) &= K(\mathbf{x}_7, \mathbf{x}_9) = \|\begin{pmatrix} 3.5 & 4 \end{pmatrix} - \begin{pmatrix} 2 & 2.1 \end{pmatrix}\|^2 \\
&= \|\begin{pmatrix} 1.5 & 1.9 \end{pmatrix}\|^2 \\
&= \begin{pmatrix} 1.5 & 1.9 \end{pmatrix} \begin{pmatrix} 1.5 \\ 1.9 \end{pmatrix} \\
&= 5.86
\end{aligned}
$$

$$
\mathbf{K} = \begin{bmatrix} 0 & 5.86 & 1.46 & 4.64 \\ 5.86 & 0 & 10 & 1.46 \\ 1.46 & 10 & 0 & 5.86 \\ 4.64 & 1.46 & 5.86 & 0 \end{bmatrix} \tag{11}
$$

(b) Find the first principal component.
By Equation 7.35 (in text)

$$
\mathbf{Kc} = \eta_1 \mathbf{c} \tag{12}
$$

where $\mathbf{K}$ is the centered kernel matrix. The centered kernel matrix is given by

$$
\begin{aligned}
\hat{\mathbf{K}} &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}_{n\times n}\right) \mathbf{K} \left(\mathbf{I} - \frac{1}{n}\mathbf{1}_{n\times n}\right) \\
&= \begin{bmatrix} -2.32 & 2.2 & -2.2 & 2.32 \\ 2.2 & -5 & 5 & -2.2 \\ -2.2 & 5 & -5 & 2.2 \\ 2.32 & -2.2 & 2.2 & -2.32 \end{bmatrix}
\end{aligned} \tag{13}
$$

The eigenvalues of the centered kernel matrix are

$$
\begin{aligned}
\lambda_1 &= -12.4719 && (14) \\
\lambda_2 &= -2.1681 && (15) \\
\lambda_3 &= -0.4752 && (16) \\
\lambda_4 &= 15.1152 && (17) \\
& && (18)
\end{aligned}
$$

Thus, we have $\eta_1 = -12.4719$. The eigenvector corresponding to this eigenvalue gives $\mathbf{u}_1$, the first kernel principal component

$$
\mathbf{u}_1 = \begin{pmatrix} 0.3463 \\ -0.6165 \\ 0.6165 \\ -0.3463 \end{pmatrix} \tag{19}
$$

3. Given the two points $\mathbf{x}_1 = (1,2)^T$, and $\mathbf{x}_2 = (2,1)^T$, use the kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^2$$

to find the kernel principal component, by solving the equation $\mathbf{Kc} = \eta_1 \mathbf{c}$.

Solution:

$$K(\mathbf{x}_1, \mathbf{x}_1) = \left[\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right]^2 = 5^2 = 25 \tag{20}$$

$$K(\mathbf{x}_2, \mathbf{x}_2) = \left[\begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right]^2 = 5^2 = 25 \tag{21}$$

$$K(\mathbf{x}_1, \mathbf{x}_2) = K(\mathbf{x}_2, \mathbf{x}_1) = \left[\begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right]^2 = 4^2 = 16 \tag{22}$$

Combining (17), (18), and (19)

$$\mathbf{K} = \begin{bmatrix} 25 & 16 \\ 16 & 25 \end{bmatrix} \tag{23}$$

We now center the kernel matrix

$$\begin{aligned} \hat{\mathbf{K}} &= \left(\mathbf{I} - \frac{1}{n} \mathbf{1}_{n \times n} \right) \mathbf{K} \left(\mathbf{I} - \frac{1}{n} \mathbf{1}_{n \times n} \right) \\ &= \begin{bmatrix} 4.5 & -4.5 \\ -4.5 & 4.5 \end{bmatrix} \end{aligned}$$

The eigenvalues of $\hat{\mathbf{K}}$ are

$$\lambda_1 = 0 \tag{24}$$
$$\lambda_2 = 9 \tag{25}$$

The first principal component, the eigenvector $\mathbf{c}$ (also $\mathbf{u}_1$) corresponding to η_1 (also λ_1), is

$$\mathbf{u}_1 = \begin{pmatrix} -0.7071 \\ -0.7071 \end{pmatrix} \tag{26}$$

4. (20 points) Consider the Prostate Cancer data from `http://statweb.stanford.edu/~tibs/ElemStatLearn/data.html`. Cast the linear regression problem over this dataset as one of linear programming, and solve it using any off-the-shelf LP solver. Present your results and analyze the performance of the regression.

Solution: I chose to consider linear regression in terms of minimizing the sum of all absolute values errors between actual values in the dataset and the value predicted for a given data point by linear regression

$$\sum_{i=1}^{n} |ax_i + b - y_i|$$

By introducing an error variable e_i for each instance x_i, we can eliminate the absolute value in the sum of errors, and cast the problem of linear regression as the following linear program

$$\begin{array}{lll} \text{Minimize} & \sum_{i=1}^{n} e_i & \\ \text{subject to} & Ax_i + b - y_i - e_i \leq 0 & \text{for } i = 1, 2, \ldots, n \\ & -(Ax_i + b - y_i) - e_i \leq 0 & \text{for } i = 1, 2, \ldots, n \end{array}$$

5. (20 points) Consider the South African Heart Disease data from `http://statweb.stanford.edu/~tibs/ElemStatLearn/data.html`. Cast the linear classification problem over this dataset as one of linear programming, and solve it using any off-the-shelf LP solver. Present your results and analyze the performance of the classifier thus learnt.

Solution: