Meet Barot Machine Learning: Problem Set 3: Kernel Methods

1.a. $k(x,z) = \phi(x) \cdot \phi(z)$.

It is a kernel, because if we let $\phi(x)$ be the feature vector for x, such that any word not in the document is given the value 0 and any word present in the document is 1. The dot product $\phi(x) \cdot \phi(z)$ would give the size of the intersection of the sets of words of x and z, which is equal to $K(x,z)$.

b. $\phi(\vec{x}) \cdot \phi(\vec{z}) = (1+\beta \vec{x} \cdot \vec{z})^2 - 1 = (1+\beta(x_1 z_1 + x_2 z_2))^2 - 1$

$= 1 + 2\beta(x_1 z_1 + x_2 z_2) + \beta^2(x_1 z_1 + x_2 z_2)^2 - 1$

$= \beta^2(x_1^2 z_1^2 + 2x_1 z_1 x_2 z_2 + x_2^2 z_2^2) + 2\beta(x_1 z_1 + x_2 z_2)$

$= \beta^2(\vec{x} \cdot \vec{z})^2 + 2\beta(\vec{x} \cdot \vec{z})$

$\phi(\vec{x}) = (\beta^2 \vec{x}^2, 2\beta(\vec{x}))$

$\phi(\vec{z}) = (\beta^2 \vec{z}^2, 2\beta \vec{x})$

$$\phi(\vec{x}) = \begin{pmatrix} \beta^2 x_1^2 \\ \beta^2 x_1 x_2 \\ \beta^2 x_2^2 \\ 2\beta x_1 \\ 2\beta x_2 \end{pmatrix}$$

c. $K(x,z) = \left(1 + \left(\frac{x}{\|x\|_2}\right)^T \left(\frac{z}{\|z\|_2}\right)\right)^3$

$= \left(1 + \left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)\right)^3$

Using (i.), since the functions $\frac{1}{\|x\|_2}$ and $\frac{1}{\|z\|_2}$ scale $x^T z$, and that $x^T z$ is a kernel, $\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)$ is a valid kernel. Using (ii.), and the fact that 1 is a valid kernel, and the fact that $\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)$ is a valid kernel, $1 + \left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)$ is a valid kernel. Now, since $\left(1+\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)\right)^2$ is a product of two valid kernels, using (iii.) it must be a valid kernel.

Finally, since $\left(1+\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)\right)^2$ is a valid kernel, using (iii) once more tells us that its product with $\left(1+\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)\right)$ is a valid kernel as well. Therefore, $\left(1+\left(\frac{1}{\|x\|_2}\right)\left(\frac{1}{\|z\|_2}\right)(x^T z)\right)^3$ is a valid kernel.

2\. c. Each lambda seemed to produce the same cross-validation error.

d. The default error for the SVC from sci-kit learn was 0.000999.

e. The 5 fold cross validation error for the SVC was 0.01549. The 10-fold error was 0.0069965.

f. When setting $C = .5$ and $\gamma = .01$, the 10-fold error became 0.00549. The test error was 0.000999, still, even with this change.