CS446: Machine Learning **Fall 2014**

Problem Set 5

Nikhil Unni *Handed In: November 6, 2014*

1. SVM

   (a) 1. $\mathbf{w} = [-1, 0]^T$
       $\theta = 0$
   2. $\mathbf{w} = [-0.5, 0.25]^T$
       $\theta = 0$
   3. I found the two closest positive/negative points, $[(-1.2, 1.6), +], [(2, 0), -]$, and found the slope between them, $\frac{1.6}{-3.2} = -\frac{1}{2}$, and the midpoint, $(0.4, 0.8)$, so the line with the farthest distance between the two points (the support vectors), has a slope of 2 with a point $(0.4, 0.8)$, giving the line $y = 2x$, which gives $w = [-2, 1]^T, \theta = 0$.

      Then, I just minimized w by halving it repeatedly, until I got $w = [-0.5, 0.25]$. This w gave $y(w^T x + \theta) = 1$ for both support vectors, so I know this is the smallest value of w I can get.

   (b) 1. $I = \{1, 6\}$
   2. $\alpha = \{\frac{5}{32}, \frac{5}{32}\}$
   3. Objective function value $= \frac{5}{32}$.

   (c) FINISH ME LATER. C represents how much the SVM should avoid misclassifications. In general, C controls the relative importance of maximizing the margin. For $C = \infty$, we obtain our original hyperplane that we found in (a)-2. For $C = 1$, we get a larger margin, with a higher chance of misclassification. The support vectors for $C = 1$ can now be inside the margins. For $C = 0$ has an even wider margin, with even larger misclassification. (FINISH ME LATER)

2. Kernels

(a)

```
1. Initialize α to 0⃗ of length n, where n is the number of examples.
2. Initialize θ to 0.
3. While still making mistakes (terminate after long string of successes):
4.     For each training example (x, y):
5.         if y[(Σ_{i=1}^{n} α_i y_i ⟨x_i, x⟩) + θ] < 0: (⟨a, b⟩ representing the inner product)
6.             α_i ← α_i + 1 (where i is the index of the current example (x, y))
7.             θ ← θ + y
```

(b)

$$K(x,z) = \alpha K_1(x,z) + \beta K_2(x,z)$$

Since K_1 and K_2 are both valid kernel functions, they can be represented as the dot product of two feature maps, ϕ_1 and ϕ_2 such that

$$K(x,z) = \alpha\langle\phi_1(x), \phi_1(z)\rangle + \beta\langle\phi_2(x), \phi_2(z)\rangle$$

And also such that

$$\phi_1(x) = [\phi_1(x)_1, \phi_1(x)_2, \ldots, \phi_1(x)_M]$$

$$\phi_2(x) = [\phi_2(x)_1, \phi_2(x)_2, \ldots, \phi_2(x)_N]$$

Where M and N the size of the vectors $\phi_1(x)$ and $\phi_2(x)$ respectively. Then, we can represent $K(x,z)$ by using the definition of ϕ_1 and ϕ_2 and expand out the inner products

$$K(x,z) = \alpha\sum_{i=1}^{M}\phi_1(x)_i\phi_1(z)_i + \beta\sum_{j=1}^{N}\phi_2(x)_j\phi_2(z)_j$$

Then we put α and β inside the summations as follows

$$K(x,z) = \sum_{i=1}^{M}(\sqrt{\alpha}\phi_1(x)_i)(\sqrt{\alpha}\phi_1(z)_i) + \sum_{j=1}^{N}(\sqrt{\beta}\phi_2(x)_j)(\sqrt{\beta}\phi_2(z)_j)$$

Suppose we had a feature map like :

$$\phi(x) = [\sqrt{\alpha}\phi_1(x)_1, \sqrt{\alpha}\phi_1(x)_2, \ldots, \sqrt{\alpha}\phi_1(x)_M, \sqrt{\beta}\phi_2(x)_1, \ldots, \sqrt{\beta}\phi_2(x)_N]$$

of dimension $N + M$.
Then, $\langle\phi(x), \phi(z)\rangle = \alpha\phi_1(x)_1\phi_1(z)_1 + \alpha\phi_1(x)_2\phi_1(z)_2 + \ldots \alpha\phi_1(x)_M\phi_1(z)_M + \beta\phi_2(x)_1\phi_2(z)_1 + \ldots + \beta\phi_2(x)_N\phi_2(z)_N$
Or

$$\langle\phi(x), \phi(z)\rangle = \sum_{i=1}^{M}(\sqrt{\alpha}\phi_1(x)_i)(\sqrt{\alpha}\phi_1(z)_i) + \sum_{j=1}^{N}(\sqrt{\beta}\phi_2(x)_j)(\sqrt{\beta}\phi_2(z)_j)$$

$$\langle\phi(x), \phi(z)\rangle = K(x,z)$$

Because $K(x,z)$ is an inner product of our new feature map, it is a valid kernel for all valid kernels K_1 and K_2 and all positive α and β.

(c) Before proving that $K(x,z)$ is a valid kernel, I'll prove one more property of kernels.
Note: I'm going to reuse K, x and z for this proof, they're not necessarily the same K, x and z we were given in the problem, sorry for the slight abuse of notation!

1. $K(x,z) = K_1(x,z)K_2(x,z)$, for all valid kernels K_1 and K_2

$$K(x,z) = (\langle\phi_1(x), \phi_1(z)\rangle) * (\langle\phi_2(x), \phi_2(z))$$

$$K(x,z) = (\sum_{i=1}^{M} \phi_1(x)_i\phi_1(z)_i)(\sum_{j=1}^{N} \phi_2(x)_j\phi_2(z)_j)$$

$$K(x,z) = \sum_{i=1}^{M}\sum_{j=1}^{N} \phi_1(x)_i\phi_1(z)_i\phi_2(x)_j\phi_2(z)_j$$

Let $\phi(x)_{ij} = \phi_1(x)_i * \phi_2(x)_j$ and $\phi(z)_{ij} = \phi_1(z)_i * \phi_2(z)_j$. Then:

$$K(x,z) = \sum_{i=1}^{M}\sum_{j=1}^{N} \phi(x)_{ij}\phi(z)_{ij}$$

$$K(x,z) = \langle\phi(x), \phi(z)\rangle$$

Where the dimension of ϕ is $M * N$.
Because K is the inner product of the feature map of x and z, it's a valid kernel.

Now that we have that, we can continue with the original proof.

First of all, $K_1(x,z) = x^Tz$ is clearly a valid kernel, where the feature map $\phi_1(x)$ is just the identity feature map, so that $K_1 = \langle\phi_1(x), \phi_1(z)\rangle$.

Next, because of my proof right above, $(x^Tz)(x^Tz) = (x^Tz)^2$ is a valid kernel too, since it's just the product of two valid kernels. By the same logic, $(x^Tz)(x^Tz)^2 = (x^Tz)^3$ is also a valid kernel.

Then, by the proof from part (b), we know that the linear combination (with positive coefficients) of valid kernels is a valid kernel as well.
So $1(x^Tz)^3 + 400(x^Tz)^2$ is valid, and then $1(x^Tz)^3 + 400(x^Tz)^2 + 100x^Tz$ is a valid kernel as well. $\blacksquare$

3. Answer to problem 3