CS446: Machine Learning **Fall 2014**

Problem Set Problem set number

My name *Handed In: Hand in date*

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. $\alpha = \left\{\frac{5}{32}, \frac{5}{32}\right\}$

   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 $\vec{0}$ of length n, where n is the number of examples.

3. Answer to problem 3