CS181 Assignment 3

Joy Ming and Alisa Nguyen (15 March 2013)

1 Problem 1

1. The probability that all of the M dimensions of $x - y$ are between $-\epsilon$ and ϵ is $\boxed{(2\epsilon)^M}$.
   For each dimension i of χ, the probability that $|x_i - y_i| \leq \epsilon$ is equivalent to

   $$P(|x_i - y_i| \leq \epsilon) =$$
   $$P(-\epsilon \leq x_i - y_i \leq \epsilon) =$$
   $$P(-\epsilon - x_i \leq -y_i \leq \epsilon - x_i) =$$
   $$P(\epsilon + x_i \geq y_i \geq x_i - \epsilon) =$$
   $$P(x_i - \epsilon \leq y_i \leq \epsilon + x_i)$$

   This distribution function is equivalent to $\int_{\epsilon+x_i}^{x_i-\epsilon} f(x)dx$, where $f(x)$ is the PDF of y_i, which we know to have a uniform distribution, so $f(x) = \frac{1}{b-a} = 1$. Thus, we get:

   $$\int_{\epsilon+x_i}^{x_i-\epsilon} 1dx =$$
   $$\epsilon + x_i - (x_i - \epsilon) = 2\epsilon$$

   Because we want to know the probability that all of the M dimensions of $x - y$ are between $-\epsilon$ and ϵ, we simply take $\Pi_{i=1}^{M} P(|x_i - y_i| \leq \epsilon) = (2\epsilon)^M$.

2. The probability of $max_m|x_m - y_m| \leq \epsilon$ is at most p

3. If x is any point in χ, and y is a point in χ drawn randomly from a uniform distribution on χ, then the probabilty that $||x - y|| \leq \epsilon$ is also at most p

4. Lowerbound on number N of points needed to guarantee

5. We can conclude that the effectiveness of the hierarchical agglomerativ clustering algorithm in high dimensional spaces

2 Problem 2

1. Given a prior distribution $Pr(\theta)$ and likelihood $Pr(D|\theta)$, the predictive distribution $Pr(x|D)$ for a new datum,

   (a) ML: $Pr(x|D) = Pr(x|\theta) = \arg\max_{\theta}(\ln(Pr(D|\theta)))$

   (b) MAP: $Pr(x|D) = Pr(x|D) = Pr(x|\theta) = \arg\max_{\theta}(\ln(P(D|\theta)P(\theta)))$

   (c) FB: $Pr(x|D) = \int \theta P(\theta|D)d\theta$

2. MAP can be considered "more Bayesian" than ML because it takes into account the distribution of θ instead of assuming same weight or uniformity.

3. One advantage the MAP method enjoys over the ML method

4. The Beta distribution is the conjugate prior of the Bernoulli.

5. Under the ML approach

3 Problem 3

1. The K -means clustering objective is to minimize the sum of squared distances between prototype and data.

2. PCA relates to K-means

4 Problem 4