CS181 Assignment 3

Joy Ming and Alisa Nguyen (15 March 2013)

# 1 Problem 1

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

$$\begin{gathered} P(|x_i - y_i| \leq \epsilon) = \\ P(-\epsilon \leq x_i - y_i \leq \epsilon) = \\ P(-\epsilon - x_i \leq -y_i \leq \epsilon - x_i) = \\ P(\epsilon + x_i \geq y_i \geq x_i - \epsilon) = \\ P(x_i - \epsilon \leq y_i \leq \epsilon + x_i) \end{gathered}$$

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:

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

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

b. The probability of $max_m|x_m - y_m| \leq \epsilon$ is at most $\rho$ because as shown in (a), $\rho$ does not depend on $x_i$ and thus holds for all $x_i$. In addition, logically, if $x$ is the center point, the average distance from it to any other point $y$ is at most $\frac{1}{2}$ for any one dimension. As $x$ moves farther and farther away from the center, the average distance increases so that it becomes at most 1 in any one dimension. So, if $x$ is not in the center $max_m|x_m - y_m|$ grows and is less likely to be less than $\epsilon$, decreasing that probability so that it is less than $\rho$.

c. We will show that $||x - y|| \geq max_m|x_m - y_m|$.

$$\begin{gathered} ||x - y|| = \sqrt{\Sigma_{m=1}^{M}(x_m - y_m)^2} \\ \sqrt{\Sigma_{m=1}^{M}(x_m - y_m)^2} \geq max_m|x_m - y_m| \\ \Sigma_{m=1}^{M}(x_m - y_m)^2 \geq (max_m|x_m - y_m|)^2 \end{gathered}$$

This is true because the left side of the inequality includes the right side in its sum. $||x - y||$ is the total Euclidean distance between two points whereas $max_m|x_m - y_m|$ is only the distance between one dimension of two points. The left side must be larger.

If $x$ is any point in $\chi$, and $y$ is a point in $\chi$ drawn randomly from a uniform distribution on $\chi$, then the probabilty that $||x-y|| \leq \epsilon$ is also at most $p$ because $||x-y||$ is greater than or equal to $max_m|x_m - y_m|$, making it less likely to be less than $\epsilon$ and thus giving it a probability lower than $\rho$ of being less than $\epsilon$.

d. Lowerbound on number $N$ of points needed to guarantee

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

## 2 Problem 2

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

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

c One advantage the MAP method enjoys over the ML method

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

e Under the ML approach

## 3 Problem 3

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

b PCA relates to $K$-means

## 4 Problem 4