CS181 ASSIGNMENT 3

ASHOK CUTKOSKY AND TONY FENG

Problem 1.

  a. This is the probability that y lies in a box of sidelength 2ϵ centered at x. This is just the ratio of the area of this box to the are of the whole box: $p = 8\epsilon^3$.
  b. This probability is the ratio of the volume of the intersection of a box of sidelength 2ϵ centered at x with the box of sidelength 1 with the volume of the box of sidelength 1. The volume of this intersection is at most the volume of the box of sidength 2ϵ, so the ratio is at most $8\epsilon^3 = p$.
  c.
$$||x - y|| = \sqrt{\sum(x_i - y_i)^2} \geq \sqrt{\max(x_i - y_i)^2} = \max|x_i - y_i|$$

  so that $||x - y|| \geq \max|x_i - y_i|$. The probability of being within distance ϵ of x is the volume of the ball of radius ϵ about x. Since distance under this euclidean metric is at most distance under the previous metric, this ball has volume at most the volume of the ball under the previous metric, which is at most p.
  d. Let p be the probability that for a given point x, a randomly drawn point y will be within ϵ of x. Then the probability that for a given point x a set of N independently randomly drawn points y will all NOT be within ϵ of x is $(1 - p)^N$. Thus the probability that the nearest neighbor of x is within ϵ of x is $1 - (1 - p)^N$.

  Now p is just the intersection of the ball of radius ϵ about x and the box of sidelength 1. The ball of radius ϵ about x contains all the points of distance ϵ from x along each of the m principle axes, and so also contains the polyhedron given by the convex closure of these points. An M-dimensional pyramid with base area A and height h has volume Ah/M. Thus if A_M is the volume of this polyhedron in M-dimensions, we have $A_M = 2\epsilon A_{M-1}/M$, with $A_1 = 2\epsilon$. Thus $A_M = 2^M\epsilon^M/M!$.

  The volume of the intersection of the ball of radius ϵ and the box is minimized when x is in a corner of the box, in which case the volume is given by the volume of the ball of radius ϵ divided by 2^M. Thus $p \geq \epsilon^M/M!$. Thus the probability that a nearest neighbor of x is witnih ϵ of x is $1 - (1 - p)^N \geq 1 - (1 - \epsilon^M/M!)^N$ and so $N \geq \log\delta/\log(1 - \epsilon^M/M!)$.
  e. Note that $\log(1 - \epsilon^M/M!)$ approaches 0 very rapidly as M increases so that the lower bound on N increases very rapidly. Intuitively, this says that points become very far apart very quickly as the dimension increases and so HAC, which relies on distances between points, might become less effective as all these distances become very large.

Problem 2.

  a. Let $\theta_{ML} = \text{argmax}_\theta(Pr(D|\theta))$, $\theta_{MAP} = \text{argmax}_\theta(Pr(D|\theta)Pr(\theta))$. Also, assign

$$Pr(\theta|D) = \frac{Pr(D|\theta)Pr(\theta)}{\int_\Theta Pr(D\theta)Pr(\theta)d\theta}$$

Then

$$Pr_{ML}(x|D) = Pr(x|\theta_{ML}$$
$$Pr_{MAP}(x|D) = Pr(x|\theta_{MAP})$$
$$Pr_{FB}(x|D) = \int_{\Theta} Pr(x|\theta)Pr(\theta|D)d\theta$$

b. The MAP method can be considered more basian because it makes more use of priors.
c. MAP enjoys the advantage of incorporating more knowledge about the world than the ML approach so that it is more resistant to overfitting. MAP enjoys the advantage of being computationally more tractable than the FB approach.
d.
e. The Beta distributio is useful because it is a conjugate distribution - $p(\theta)$ has the same functional form as $p(\theta|x)$. This causes $p(\theta|x)$ to have a very simple form.
f. In Fall 2011, the Harvad football team won 9 games and lost once. Using the the ML model, the probability of the team winning any particular game is then just $\frac{9}{10}$, and so the probability of winning the next game is also $\frac{9}{10}$.

   Now we calculate using the MAP model. Let's adopt the $\beta(5,3)$ prior. Then following the lecture notes,

$$\text{argmax}_\theta P(\theta|D) = \frac{5+9-1}{3+5+10-1} = \frac{13}{17}$$

so that the probability of winning the next game is $\frac{13}{17}$.

   Now we use the FB model, again with the $\beta(5,3)$ prior. Then we have

$$\text{argmax}_\theta P(\theta|D) = \int_0^1 \theta p(\theta|D)d\theta = \frac{5+9-1}{3+5+10-2} = \frac{13}{16}$$

so that the probability of winning the next game is $\frac{13}{17}$.

Problem 5.

a. The loss function for vectors $x_1, \cdots, x_N$ with prototypes $y_1, \cdots, y_K$ and responsibilities r_{ij} is

$$\mathcal{L} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk}|x_n - y_k|$$

Suppose we are in M-dimensional space, with $x_n = (x_{n,1}, \cdots, x_{n,m})$ and $y_k = (y_{k,1}, \cdots, y_{k,m})$. Then we have

$$\frac{\partial \mathcal{L}}{\partial y_{k,m}} = \sum_{n=1}^{N} r_{nk} \frac{2(y_{k,m} - x_{n,m})}{2|x_n - y_k|}$$

Problem 4.

a. (a)

k	Mean squared distance
1	1.97
2	1.80
3	1.81
4	1.77
5	1.52
6	1.37
7	1.42
8	1.25
9	1.31
10	1.26

(a) (b) It is difficult to tell, because the mean squared values always seem to be improving as the number of clusters increases (which should be the case). However, the improvements seem to level off somewhat at 8 clusters, so that is about as reasonable a guess as we can make just from the above data.
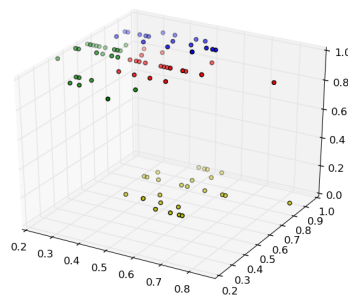
b. (a) Number of instances in each cluster by metric:

Cluster Number	Min metric	Max metric
0	73	19
1	25	23
2	1	32
3	1	26

Scatterplot of Min metric clusters:
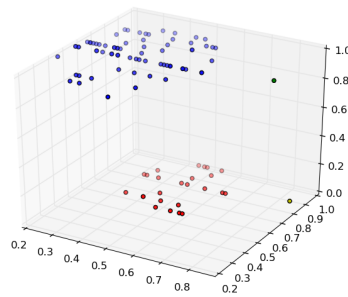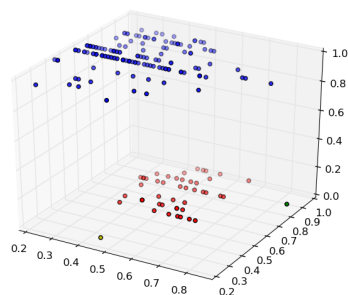

Scatterplot of Max metric clusters:

(b) Number of instances in each cluster by metric:

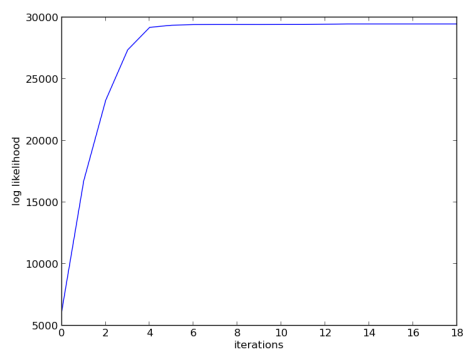| Cluster Number	mean metric	centroid metric
0	152	147
1	46	47
2	1	5
3	1	1

Scatterplot of mean metric clusters:

Scatterplot of centroid metric clusters:
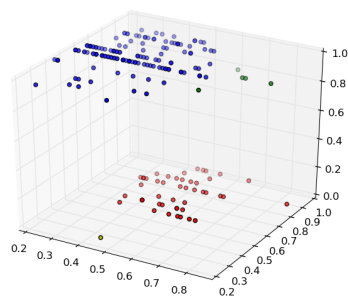
c. (a) It takes autoclass 18 iterations to converge. The cluster sizes were 239,394,363 and 4.

(b)

(c) Autoclass takes about a minute to classify 1000 instances. K-means takes about a second. Thus K-means has a significantly better run-time than Autoclass.