CS-543: Computer Vision

Instructor: D.A. Forsyth

Homework 3

Instructions

This homework contains four questions, of equal value. Prepare a web page containing written answers to two of the following questions, and email the URL to me by April 30, 2008. Please use the term “homework” in the subject line of your email, and give me the names of all participants. You may work in groups of up to three. You may write out the answers and scan them, or type them, or whatever.

Question 1: Segmentation

How should a segmenter be evaluated? your answer should discuss available methods, and assess them. **Hint:** *If you think that this is an easy question, or one with an obvious answer, keep thinking.*

Question 2: EM

We are presented with a dataset of the following form. The i’th data item, $\mathbf{X}_i$, consists of two English words, E_{i1} and E_{i2}, and two French words F_{i1} and F_{i2}. The English words are drawn from a vocabulary of 100 words, and the French words are drawn from a vocabulary of 100 words. The data item is obtained as follows: first, we draw a pair — one English word and one French word — from the probability distribution $P(\text{English}, \text{French})$; second, we draw a second pair independently from this distribution; now we “drop” the correspondence information. For example, the data item (“cat”, “dog”, “rouge”, “noir”) could be obtained by drawing (“cat”, “rouge”) then (“dog”, “noir”) or by drawing (“dog”, “rouge”) then (“cat”, “noir”). If this pair was obtained by drawing (“cat”, “rouge”), etc. then we say the pair (“cat”, “rouge”) *corresponds*.

From this dataset we wish to estimate $P(\text{English}, \text{French})$. We introduce some notation. Write $p_{kl} = P(\text{English} = \text{word } k, \text{French} = \text{word } l)$. This is a table, with $100 \times 100 - 1$ degrees of freedom. Write E_{ij} for the index of the j’th (which is 1 or 2) English word in data item i (by this, I mean which word in the vocabulary). Write F_{ij} for the index of the j’th French word in data item i.

Part 1: Show that this means that the incomplete data log-likelihood is:

$$\sum_i \log\left(p_{E_{i1},F_{i1}} p_{E_{i2},F_{i2}} + p_{E_{i1},F_{i2}} p_{E_{i2},F_{i1}}\right)$$

Part 2: Introduce a hidden variable δ_i. If E_{i1} and F_{i1} correspond, then $\delta_i = 1$ otherwise it is 0. Show that the incomplete data log-likelihood is:

$$\sum_i \left[\left(\log\left(p_{E_{i1},F_{i1}} p_{E_{i2},F_{i2}}\right)\right)\delta_i + \left(\log\left(p_{E_{i1},F_{i2}} p_{E_{i2},F_{i1}}\right)\right)(1 - \delta_i)\right]$$

Part 3: Show that the update equations for EM are the same as those you would get if you did coordinate ascent on

$$\sum_i \left[\left(\log \left(p_{E_{i1},F_{i1}} p_{E_{i2},F_{i2}} \right) \right) \delta_i + \left(\log \left(p_{E_{i1},F_{i2}} p_{E_{i2},F_{i1}} \right) \right) (1 - \delta_i) \right] - \sum_i \left(\delta_i \log \delta_i + (1 - \delta_i) \log (1 - \delta_i) \right)$$

as a function of the p_{kl} and the δ_i.

Part 4: Show that this optimization problem is convex.

Question 3: Tracking

We have a linear dynamical system with states $\mathbf{X}_i$. In particular, $\mathbf{X}_i$ is a normal random variable with mean $\mathcal{D}\mathbf{X}_{i-1}$ and covariance Σ_d. At each time step, there are *two* components to the measurement $\mathbf{Y}_i$. The components are $\mathbf{Y}_{i1}$ and $\mathbf{Y}_{i2}$. One of these two components is a normal random variable with mean $\mathcal{M}\mathbf{X}_i$ and covariance Σ_m. The other is a normal random variable with mean 0 and covariance Σ_n. We do not know which of the components is informative. $P(\mathbf{X}_1)$ is Gaussian.

Part 1: Show that, if $P(\mathbf{X}_{i-1}|\mathbf{Y}_1, ..., \mathbf{Y}_{i-1})$ is a mixture of k Gaussians, then $P(\mathbf{X}_i|\mathbf{Y}_1, ..., \mathbf{Y}_{i-1})$ is a mixture of k Gaussians.

Part 2: Show that, if $P(\mathbf{X}_i|\mathbf{Y}_1, ..., \mathbf{Y}_i)$ is a mixture of k Gaussians, then

$$P(\mathbf{X}_i|\mathbf{Y}_1, ..., \mathbf{Y}_i, \text{first component of measurement is informative})$$

is a mixture of k Gaussians.

Part 3: Show that the posterior, $P(\mathbf{X}_i|\mathbf{Y}_1, ..., \mathbf{Y}_i)$ is a mixture of Gaussians, with 2^i mixture components.

Part 4: We cannot use an exact representation of the posterior in this case, because there are too many components. Suggest a strategy for managing this problem.

Question 4: Multiple Views

In this Problem, we use the right-handed coordinate systems. A photographer took a pair of images as follows. After taking the first image, he moved the camera and took another image. The vector $\mathbf{P}$ representing a point of the first image and the vector $\mathbf{Q}$ representing a point of the second image satisfied

$$\mathbf{P}\mathcal{F}\mathbf{Q} = 0$$

where the fundamental matrix was $\mathcal{F}$ was estimated as

$$\mathcal{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 10 \\ 0 & -20 & 0 \end{pmatrix}.$$

Part 1:(5 pts)

In general, do choices of image coordinate systems affect the simple 8-point least-squares fitting estimation of the fundamental matrix? Explain why.

Part 2:(5 pts)

Calculate the epipolar line of an image point of the first image whose image coordinates are (p, q).

Part 3:(5 pts)

Find out the relative positions of the two cameras.

Part 4:(5 pts)

Find out the ratio of the focal lengths of the two cameras.

Part 5:(5 pts)

Calculate the three-dimensional coordinates of a point whose image coordinates in the first image is $(0, 0)$ and the image coordinates in the second image is $(1, 0)$ up to the scaling factor, in the coordinate system attached to the first camera.