Exam 2: Computer Vision (600.361/600.461)

Instructor: René Vidal

December 5, 2013

Part I (10 points) Answer these questions in 1-4 lines.

1. **(1 points)** Name an interactive segmentation algorithm.
2. **(2 points)** What are the main similarities and differences between K-means and EM for a mixture of Gaussians.
3. **(2 points)** What is the difference between the graph Laplacian and the symmetrized graph Laplacian?
4. **(2 points)** What is a classifier?
5. **(1 points)** What is a linear classifier?
6. **(2 points)** How are the visual words obtained in the Bag-of-Features approach to image classification?

Part II (30 points) Solve the following problems.

1. **(15 points) Spectral Clustering.** Let $G = (V, E)$ be a weighted undirected graph with vertex set $V = \{1, 2, \dots, N\}$, edge set $E = \{(i, j) : i, j \in V\}$, and weights $w_{ij} = w_{ji} \geq 0$. Let $W \in \mathbb{R}^{N\times N}$ be the weighted adjacency matrix, $D \in \mathbb{R}^{N\times N}$ be the (diagonal) degree matrix with entries $d_i = \sum_{j=1}^{N} w_{ij}$, and $L = D - W \in \mathbb{R}^{N\times N}$ be the Laplacian matrix.

   (a) **(2 points)** Show that $L\mathbf{1} = \mathbf{0}$, where $\mathbf{0}$ and $\mathbf{1}$ are the vector of all zeros and ones, respectively.

   (b) **(3 points)** Let $\boldsymbol{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$. Show that
   $$\boldsymbol{x}^T L \boldsymbol{x} = \frac{1}{2}\sum_{i,j=1}^{N} w_{ij}(x_i - x_j)^2. \tag{1}$$

   (c) **(10 points)** Let $V_1, \dots, V_K$ be a partition of V, that is $V = V_1 \cup \dots \cup V_K$, and $V_i \cap V_j = \emptyset$, $i \neq j = 1, \dots, K$. Let $H \in \mathbb{R}^{N\times K}$ be defined as $h_{ij} = \begin{cases} 1 & \text{if } i \in V_j \\ 0 & \text{else} \end{cases}$. Show that $H^\top D H = I$ and that
   $$\text{Cut}(V_1, \cdots, V_K) = \sum_{k=1}^{K} \text{Cut}(V_k, \bar{V}_k) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i\in V_k, j\in \bar{V}_k} w_{ij} = \text{trace}(H^\top L H). \tag{2}$$

2. **(5 points) SVM Classification.** Given a training set $\{(\boldsymbol{x}_j, y_j)\}_{j=1}^{N}$, where $\boldsymbol{x}_j \in \mathbb{R}^D$ is a data point and $y_j \in \{-1, 1\}$ is its label, a linear SVM classifier is trained by solving the following optimization problem
   $$\min_{\boldsymbol{w},\beta} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{such that} \quad y_j(\boldsymbol{w}^\top \boldsymbol{x}_j + \beta) \geq 1 \quad j = 1, \dots, N. \tag{3}$$
   Consider the Lagrangian function $\mathcal{L} = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{j=1}^{n} \alpha_j(1 - y_j(\boldsymbol{w}^\top \boldsymbol{x}_j + \beta))$, where α_i are the Lagrange multipliers. Show that the optimal separating hyperplane is given by
   $$\boldsymbol{w} = \sum_{j=1}^{N} \alpha_j y_j \boldsymbol{x}_j. \tag{4}$$

3. **(10 points) Object Detection**

   Suppose you are given images of different table settings, such as the one illustrated in the figure. Suppose you have trained a spoon detector and a knife detector. Describe in detail the design of a knife-spoon detector that fires when a knife and a spoon appear next to each other and parallel to each other, e.g., as shown in the figure. In particular, describe how you would represent the pose of each object, and how you would search efficiently for both objects in multiple locations and scales.