Solutions

Problem 1: Recall (2pts) (Answer in one sentence only.)

What is the objective criterion of spectral clustering (i.e. what value is the clustering solution trying to optimize, and what is considered optimal)?

Spectral clustering seeks to (approximately) minimize the graph cut or normalized graph cut criterion (in other words, the total weight of between-cluster edges, or the ratio of between-cluster to within-cluster edges).

Problem 2: Work (8 pts) (Show all derivations/work and explain.) Fisher Linear Discriminants, like PCA, can be expressed as a standard eigenproblem:

$$S_W^{-1} S_B w = \lambda w \ , \tag{1}$$

where w is our projection, S_W is the within-class scatter matrix, and S_B is the between-class scatter matrix. Under certain circumstances, we can equivalently solve:

$$w^* = S_W^{-1}(m_1 - m_2) \ , \tag{2}$$

where m_1 and m_2 are the means of our two classes.

First, under what circumstances would we be *unable* to rewrite the problem in the form seen in equation 2?

We would be unable to use this form if the problem is multiclass (that is, more than 2 classes). We would also be unable to use this form if W_S is nonsingular (technically, we would not be able to use the first form either, but we would be able to solve the equivalent general eigenproblem $S_B w = \lambda S_W w$).

Second, if we *can* use the second form of the FLD equation, what are the advantages of doing so?

The form seen in equation 2 only requires us to compute a matrix inverse, followed by matrix-vector multiplication, while equation one requires us to compute a matrix inverse, matrix multiplication, and, worst of all, an eigendecomposition. This makes the equation 2 version significantly less expensive to compute (specifically, a matrix inverse or matrix multiplication can be solved in $\mathbf{O}(n^{2.373})$, while eigendecomposition is $\mathbf{O}(n^3)$ at best, not to mention approximate, with additional complexity costs dependent on the degree of accuracy required).

Finally, even if the circumstances of our problem do allow us to compute an FLD solution using equation 2, why might we choose to use the method described in equation 1 instead?

Next page.

The chief advantage of the eigendecomposition approach is that it allows us to learn a multidimensional representation of our data by computing multiple eigenvectors corresponding to different eigenvalues. For many more complex distributions, a single linear projection is insufficient to sufficiently discriminate the data, and learning multiple orthogonal projections (as in PCA, but, because this is FLD, still specifically formulated for discrimination) can yield a more powerful dimension-reduced representation.

There was something of an error in this question though, and it would have been better phrased as "What are the advantages of the eigendecomposition method?" The error related to the fact that the matrix W_b (and by extension the matrix $W_S^{-1}W_B$) has rank $c - 1$, where c is the number of classes in the problem. Because of this, the eigenproblem described in equation 1 will only have $c - 1$ nonzero eigenvalues, and thus only $c - 1$ useful eigenvectors, so in fact if the problem is 2-class, and we can use equation 2, there is no reason to do eigendecomposition because eigendecomposition will only yield a single usable projection anyway. Quizzes were graded with this issue in mind.