Problem Set IV, due Friday Nov. 20

Problem 1. You have been provded with a set of back and white photographs. For each photograph divide the piture in 8×8 pixel cells. Each cell defines a 64 dimensional vector. Compute the PCA of the collection of all the cells of al the images. Produce a figure showing the eigenvectors, each represented as an 8×8 image, and sorted by eigenvalue. For each image take each cell of the image and project it onto the k principle eigenvectors and then reconstruct the cell form this projection. Show the reconstructed images for various values of k.

Compare the cell eigenvectors to the basis vectors of the discrete Fourier transform used in jpeg

http://en.wikipedia.org/wiki/File:Dctjpeg.png

and the order in which they are assigned importance in jpeg

http://en.wikipedia.org/wiki/File:JPEG_ZigZag.svg

The similarly of the image patch eigenvectors to the Fourier basis is not an accident. It arises from translation invariance of image statistics — the probability of an image is the same as the probability of a translation of the image. There is a general theorem that the eigenvectors of any "circulant" matrix form a Fourier basis. See

http://en.wikipedia.org/wiki/Circulant_matrix

Problem 2. In this problem we consider the relationship between PCA and L_2 regularization. More specificaly, consider a sample of n training pairs (x_1, y_1), $\ldots$, (x_n, y_n) with $x \in \mathcal{X}$ and $y_i \in \{-1, 1\}$ and consider a feature map $\Phi : \mathcal{X} \to \mathbb{R}^d$ with $d > n$. In particular we compare the following learning algorithms

A. L_2-regularized logistic regression.

$$w^* = argmin_w \sum_{t=1}^{n} (1 + e^{-m_t}) + \frac{1}{2}||w||^2$$

B. First use PCA on the vectors $\Phi(x_1)$, $\ldots$, $\Phi(x_n)$ to construct a reduced dimensionality feature map $\Phi' : \mathcal{X} \rightarrow \mathbb{R}^{d'}$ with $d' << n$. Then apply *unregularized* logistic regression under this reduced feature map.

$$w^* = argmin_w \sum_{t=1}^{n} (1 + e^{-m_t})$$

Explan intuitively why you might expect these learning algorithms to produce similar weight vectors (the lower dimensional vectors form a subspace of the higher dimensional vectors).

Compare these two algorithms (in terms of generalization performance) on the data set provided for various values of k.