Homework 1: Spring 2020. EN. 601.783: Vision as Bayesian Inference. 
Due Friday (midnight) 6 March 2020.

A.L. Yuille:

Abstract

The homework assignment is based on the first four weeks of the course. These questions should have short answers. A few sentences, some mathematics, but not long mathematical derivations. Please submit your homework via Gradescope. Pdf files are strongly preferred. Hand-written file should be scanned and submitted electronically as pdfs'.

Lecture 2: 40 points

1. What are orthogonal basis functions? How can an input image patch be expressed as a combination of orthogonal basis functions?

2. Give two examples of orthogonal basis functions.

3. Give a method for estimating a set of basis vectors given a training set of images. What form do these basis functions take if the image is shift-invariant?

4. How can we represent images in terms of a linear combination of over-complete basis functions by imposing a sparsity constraint?

5. What is the miracle of sparsity? Describe $L_1$ sparsity and show, for a simple example, how it results in a sparse representation.

6. Discuss the relative advantages of Principal Component Analysis and Sparse Coding for face recognition.

Lecture 3: 30 points

1. What is the k-means algorithm? What are the means, the assignment variable, and $k$? What are its convergence properties? What are the advantages of k-means++?

2. How can k-means be used to learn a set of dictionary elements for image patches?

3. What is a mixture of Gaussian distribution? And how does k-means relate to a mixture of Gaussian distributions?

4. What are mini-epitones? How do they deal with shift-invariance? What algorithm is used to learn them? How well can they represent images?

Lecture 4: 30 points

1. What is the Expectation-Maximization (EM) algorithm? How can EM be applied to learning a mixture of Gaussian distributions? Describe why the EM algorithm converges.

2. What are super-pixels? Briefly describe the SLIC algorithm. What are the advantages of representing an image in terms of super-pixels?
Lecture 5: 40 points

1. What is the Mumford and Shah model for image segmentation?
2. What is convexity? What is the steepest descent algorithm? Why is convexity important for steepest descent?
3. What is the Rudin-Osher-Fatemi, or total variation, model? Why is it more practically useful than the weak membrane model? Why is it less effective than dictionary methods for denoising images?
4. What is variational bounding and CCCP? How do they compare to steepest descent? How do they guarantee that each iteration decreases the cost?
5. What forms do the histograms of derivative operators of images normally take?

Lecture 6: 40 points

1. Briefly describe statistical edge detection. What cues are used for detecting edges?
2. How is this formulated as a classification learning task using groundtruth data? What image features can be used?
3. What is the log-likelihood tests? What happens if the image feature is a single cue like the derivative of the image intensity?
4. Why is a statistical approach useful if edge detection requires using multiple cues?
5. How does this relate to Bayes Decision Theory? Which is worse for edge detection, false positives or false negatives? How do the prior and loss function contribute to the threshold?
6. Give an example of another visual task that can the same approach be applied to?

Lecture 7: 40 points

1. How does regression relate to decision theory?
2. What is binary, or logistic, regression? How can this be be learnt from data? Does this require solving a convex or non-convex optimization problem? What algorithms can be used to solve it?
3. How does logistic regression relate to multi-level perceptrons and deep networks? What are the main similarities and the biggest differences?

Lecture 8: 40 points

1. Explain how the output of a deep network can be expressed as a composition of operations at different layers.
2. What is the loss function? Is it usually a convex or concave function? Can the loss function contain terms at different levels of the hierarchy?
3. Why is it important the outputs are differentiable functions of the weights. What is steepest descent? What is stochastic gradient descent? Explain how the derivatives with respect to the weights can be computed by backpropagation.
4. What are convolutional layers? What is pooling? What are the advantages of ReLu as compared to sigmoid functions for the non-linearities?
5. Give three examples of different types of deep networks.