Applied Information Theory

http://www.cs.umass.edu/∼elm/graphics/

Test 2: Review Sheet

I will be in my office on Monday from 3:30-6:00 to answer questions about the material.
You should know the following material:

1. Understand the basics of the four semi-parametric models covered in class. These are: product distributions (or factorial models), ICA, tree-graph models, and TCA.
   - Compare their advantages and disadvantages to parametric models.
   - Know the Chow-Liu algorithm (This is building a tree graph over random variables by constructing a maximum spanning tree, using the mutual information between nodes as the weight between edges.)
2. Know the Pythagorean theorem for KL-divergences. Understand its relationship to the traditional Pythagorean Theorem. Use this to show that the best product distribution approximation to a joint distribution is the product of the marginal distributions.
3. Argue that the pixel Y next to a given pixel X does not have the greatest mutual information with Y, of all pixels within the image. Make the same argument for the furthest pixel from X in the image.
4. Explain alignment of medical images by maximization of mutual information. Explain why alignment by minimizing image differences or maximizing correlation may not work for certain pairs of images.
5. Give 2 ways to estimate the differential entropy of a distribution from a sample. (Parzen windows followed by Monte Carlo and Vasicek estimator.)
6. What are the two important criteria for choosing features for object recognition (and other classification problems)? Answer: Features should be informative (have high mutual information with class label) and be non-redundant (two features shouldn't give you the exact same information).
7. Understand the joint alignment algorithm "congealing" which jointly aligns images by minimizing the entropies through the pixels. Why use entropy as a criterion instead of say, the variance? (Answer: the "true" distributions of pixel values, when aligned, might be bimodal, which would have low entropy, but not low variance. Example: MRI images.) Understand the difference between congealing and alignment by maximization of mutual information.
8. Understand the probability integral transform.
9. Know the definition of conditional mutual information. Know the definition of information gain (this is the difference between the joint mutual information $(X,Y;C) and mutual information$(Y;C).
10. Understand the difference between a code that can detect a particular type of error and a code that can detect and correct the error.
11. Understand why the Vasicek entropy estimator works.
12. Show for an empirical distribution q(x) over a discrete alphabet and a family of distributions $p(x;\theta)$ parameterized by θ that the θ that maximizes the likelihood of the data from the empirical distribution is the same as the one that minimizes the KL-divergence.