Revision List for Machine Learning and Pattern Recognition (MLPR): 2013/4

Instructor: Charles Sutton
csutton@inf.ed.ac.uk

April 15, 2014

1 Revision List

This provides a list of things that you should study in preparation for the exams. The course slides and the Murphy course text should provide the resources you need for this course. Examinable readings for each topic are given on the "Lectures" page of the Web site.

The exam will have a similar structure and similar types of questions to previous years. I highly recommend that you study the exam papers from past years in order to prepare. When you do this, however, you should be aware that there are a few topics that have been in covered in the past that we did not cover this year, for example, principal components analysis and Gaussian processes. Those topics will not appear on the exam this year.

Any topic covered in the lectures is examinable unless the slides specifically state otherwise. The list below is a list of the topics which I feel are the most important to revise:

1. The necessity of prior information, knowledge or assumption. No free lunch.
2. Different types of data: categorical, numeric, ordinal. 1 of m encoding. Why 1 of m is done. When it shouldn't be used.
3. Event space, random variables, probability mass and probability density. Cumulative probability density.
4. Distributions listed in the lecture. You should know the density function for the multivariate Gaussian distribution, the Bernoulli distribution, the Binomial Distribution and the Multinomial distribution.
5. Models. The likelihood and the log likelihood. Comparing hypotheses by maximum likelihood. Maximum likelihood for Bernoulli distribution (derivation and result).
6. Naive Bayes. Form of naive Bayes model. Conditional independence assumption for naive Bayes model: what it means and how it is written. Derivation of naive Bayes for a binary features (same as maximum likelihood for Bernoulli/ conditional Bernoulli).
7. The class conditional Gaussian classifier. Form of the conditional distribution $p(y|x)$ and the decision boundary.
8. The prior distribution. Maximum a posteriori (MAP) parameter estimation. Maximum log posterior is maximum likelihood plus penalty term (regularisation).
9. Marginalisation as integration over unknowns. Inference as marginalisation. Marginal likelihood.
10. Bayesian model comparison using the marginal likelihood.

11. Conjugacy and conjugate priors. Given a conjugate prior and likelihood, you should know how to write down the posterior.

12. Form of the density function for multidimensional Gaussian. Estimating the mean and covariance.

13. The regression model. Linear regression. Using features to do nonlinear regression.

14. Logistic regression model. The neural network model.

15. Gradient ascent. Batch versus online gradient ascent.

16. What does the Laplace approximation do. Laplace approximation by curvature matching. Computing the Hessian at the maximum. How to set the Gaussian distribution given the Hessian and maximum.

17. The KL divergence, and characteristics. Variational approximation as minimizing KL divergence between approximate posterior and true posterior. Compare benefits of Laplace v Variational approaches.

18. Importance sampling, rejection sampling, and slice sampling. You should understand these well enough that you could write down the equations necessary to apply importance sampling or rejection sampling to a new distribution that I specify.

19. MCMC sampling. Ergodicity. Detailed Balance. Metropolis Hastings, Gibbs sampling. Hamiltonian Monte Carlo as using momentum (ball rolling) combined with Gibbs sampling the momentum (giving the ball a push). Draw pictures of each in action.

20. Mixture models, as used both for clustering and as a prior distribution (mixture of conjugate priors)

21. The general form of EM algorithm. The EM algorithm for Gaussian mixtures