Exam Questions for the Data Analysis and Bayesian Inference of 851.305 "Computational Mathematics and Bioinformatics"

Peter Sykacek
Department of Biotechnology
Bioinfromatics Research Group
BOKU University
peter.sykacek@boku.ac.at

May 23, 2008

1 Allowed Material

The following definitions of probability distributions may be used during the exam. You may in addition use the lecture notes which were handed out during my presentation, as long as they do not contain pre worked solutions.

Since this is a rather generous exam setup, we do not tolerate any additional "aid" during the exam like spying on your neighbour, chatting with your neighbour or using pre-worked examples.

Univariate and multivariate Gaussian distribution:

$$p(x|\mu, \lambda) = (2\pi)^{-\frac{1}{2}}|\lambda|^{\frac{1}{2}} \exp(-0.5\lambda(x-\mu)^2)$$

$$p(\boldsymbol{x}|\boldsymbol{\mu}\boldsymbol{\Lambda}) = (2\pi)^{-\frac{d}{2}}|\boldsymbol{\Lambda}|^{\frac{1}{2}} \exp(-0.5(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu}))$$

Gamma distribution:

$$p(\lambda|g, h) = \frac{h^g}{\Gamma(g)}|\lambda|^{(g-1)} \exp(-h\lambda)$$

Multinomial one distribution:

$$P(I|\pi) = \prod_{k=1}^{K} \pi_k^{\delta(I=k)}$$

Dirichlet Distribution:

$$p(\pi) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$$

Note! for $K = 2$ we get the Bernoulli distribution as special case of the Multinomial-one distribution:

$$P(I|\pi) = \pi^{\delta(I=1)}(1-\pi)^{\delta(I=2)}$$

and the Beta distribution as a special case of the Dirichlet distribution:

$$p(\pi) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\pi^{\alpha_1-1}(1-\pi)^{\alpha_2-1}$$

Student-t distribution:

$$p(\mu|\theta,\kappa,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}|\kappa|^{0.5}(\nu\pi)^{-0.5}\left(1+\frac{(\mu-\theta)^2\kappa}{\nu}\right)^{-\frac{\nu+1}{2}}$$

2 Questions

The following questions are typical for what you will be asked during the written exam in a slightly modified fashion. Note that the exam in 2008 considers questions 2.1) to 2.5). The problems which concern Bayesian inference are not part of this years exam.

2.1 Sherman-Morrison-Woodbury

Prove that the Sherman-Morrison-Woodbury formula is correct.

2.2 Matrix Functions

Find a compact expression for vec $\boldsymbol{x}\boldsymbol{x}^T$. Explain your choice by working out some prototypical lines for both expressions.

2.3 Avoiding Integration

Calculate the expectation of x under a truncated normal distribution:

$$\int_{x=\alpha}^{\beta} x(2\pi)^{-0.5}\lambda^{0.5}\frac{1}{\Phi(\beta;\mu,\lambda)-\Phi(\alpha;\mu,\lambda)}\exp(-0.5\lambda(x-\mu)^2)dx$$

What does Φ represent here? Hint: Think about the "normalisation property" of all density functions. Find a closed form expression for the expectation. Hints: The truncated normal distribution is only non-zero in the range from α to β. The integral is solved by finding a function $F(x)$ such that $\frac{dF(x)}{dx}$ allows expressing the above intergand. The nature of the derivative of the exponential function $\exp(x)$ is of particular importance in finding this $F(x)$.

2.4 Likelihood and Linear Regression

Derive the maximum likelihood (ML) estimates for parameters $\boldsymbol{\theta}$ and λ (precision of the Gaussian noise model) for linear regression with Gaussian noise. Find a *qualitative interpretation* of the size of λ in dependence of the prediction error. Hints: the expression of the ML estimate of λ ($\hat{\lambda}$) depends on the regression parameter $\boldsymbol{\theta}$. You may use the ML estimate $\hat{\boldsymbol{\theta}}$ in the expression of $\hat{\lambda}$. Provide an argument why this is the case!

2.5 Classification in the Sampling Paradigm

Assume a two class problem. Class 1 has prior probability π and a class conditional density which is a multivariate d-dimensional Gaussian with covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}_1$. Class 0 has a class conditional density which is a multivariate d-dimensional Gaussian with covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}_0$. Answer the following questions:

- What is the class prior of class 0?
- Show that the posterior probability is given by a logistic distribution.
- Express the regression parameter $\boldsymbol{\theta}$ in terms of the parameters of the class conditional densities and class priors.

2.6 Inferring Probabilities Observing Head or Tail

Formulate Bayes theorem and derive the posterior density function over the probabilities of observing head or tail for a possibly biased coin. Consider an uncertainty in the model class and derive the model probability $P(I|\mathcal{D})$ that a sequence of draws was generated by a biased coin with unknown head vs. tail probability or from an unbiased coin with equal head and tail probability.

2.7 Utility of Medical Diagnosis Tests

Tabulate a utility function over all choices and uncertain outcomes and derive the expected utility for the following problem: We have 2 different screening tests to reveal whether a particular disease is present or not. Test A costs 2000 Euro and delivers the correct disease status with probability 0.90. Test B costs 5000 Euro and detects the correct disease status with probability 0.99. Not detecting the disease increases the cost of later diagnosis and treatment from 10000 Euro to 50000 Euro. Without testing the probability of correctly guessing the disease status is 0.5. Which test do you chose: test A, test B or none?

2.8 Inferring the Precision of a Zero Mean Gaussian

Formulate inference for the unknown precision parameter of a zero mean Gaussian distribution. Derive an expression for the posterior density function over the precision parameter. Consider model uncertainty and derive the probability $P(I|\mathcal{D})$ that a sequence of observations was generated by a zero mean Gaussian with unknown precision or from a zero mean Gaussian with unit standard deviation.

Citations

Further information about linear algebra can be found in [7], [6] and [3]. Information about matrix differential calculus (including many pre-calculated relations) is available in [5]. A good reference for mathematical functions is in [1].

An overview about data analysis which goes far beyond the topics discussed in this lecture is provided by [2].

Among the many introductions to Bayesian inference, we refer to [8, 8]. The book by David MacKay [4] contains several chapters on data analysis and Bayesian inference. This book is in particular useful since most chapters are provided online on his homepage.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* Dover, New York, 1965.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition.* Wiley, New York, 2000.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, second edition, 1989.

[4] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, Cambridge UK, 2003.

[5] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus.* Wiley, New York, 1988. paper back in 1999.

[6] C. D. Meyer. *Matrix analysis and applied linear algebra.* SIAM, 2000.

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C++.* Cambridge University Press, Cambridge, second edition, 2002.

[8] D. S. Sivia. *Data analysis a Bayesian tutorial*. Clarendon Press, Oxford, UK, 1996.