Other Questions for 3F3: Pattern Processing

Zoubin Ghahramani
Department of Engineering
zoubin@eng.cam.ac.uk

These questions are not meant to be in the same format as exam questions, but they should help you study and understand the material.

1. Is clustering a supervised or unsupervised learning problem? What about classification?

2. Describe what we mean by *overfitting* and *underfitting* in the context of polynomial regression? How about in the context of clustering?

3. Prove that $P(X = x, Y = y) \leq P(X = x)$.

4. Prove that the entropy $H(X) \geq 0$. Describe distributions for which $H(X) = 0$.

5. For each of the distribtions in lecture 1, what is the mean, variance, and entropy?

6. If $\mathbf{x}$ is multivariate Gaussian with mean μ and covariance matrix Σ and $\mathbf{y}$ is independent of $\mathbf{x}$ and multivariate Gaussian with mean m and covariance matrix S, what is distribution of $\mathbf{z} = \mathbf{x} - \mathbf{y}$?

7. Prove that multivariate Gaussians are closed under linear transformations. That is, if $\mathbf{x}$ is a D-dimensional Gaussian, then so is $\mathbf{y} = A\mathbf{x}$ where A is a $K \times D$ matrix. Comment on what happens depending on the rank of A.

8. Show that the contours of equal probability for a multivariate Gaussian are ellipses.

9. Derive the maximum likelihood equations presented in Lecture 2 for linear regression.

10. Consider polynomial regression with parameters β. Assume a prior $p(\beta)$ is Gaussian with mean 0 and variance λI where I is the identity matrix. How does the maximum likelihood estimate of β compare to the MAP estimate, and what is the effect of λ on the MAP estimate?

11. Consider classification as described in Lecture 3, but with the following model
$$y^{(n)} = H\left(\beta^\top \tilde{\mathbf{x}}^{(n)} + \epsilon_n\right)$$
where ϵ_n is Gaussian with mean 0 and variance σ^2. Compute the probability
$$P(y^{(n)} = 1 | \tilde{\mathbf{x}}^{(n)}, \beta, \sigma)$$
in terms of the Gaussian cumulative distribution.

12. Consider the function $y = x^2 + \sin(4x) + \log(1 + x^2)$. Plot this function in Matlab or Octave for values of x in the interval $(-3, 3)$. Write an iterative algorithm for finding the value of x which minimizes y. Starting at any point x_0 this algorithm takes small steps in the direction which decreases y:
$$x_{t+1} = x_t - \eta \frac{dy}{dx_t}$$
Discuss what happens for very small η and very large η. Comment on what we might mean by the concept of *local optimum* and *global optimum*.

13. Implement the online logistic classification learning rule in Matlab/Octave and play with it to see who it works. How does the learning rate affect the algorithm.

14. Implement the K-means algorithm. Play with different ways of initializing the means, and different values of K.

15. Prove that each step of K-means decreases the step the cost function C described in Lecture 4. ‘

16. How would you choose K?

17. What happens if you run K means with $K = 2$ on data from two clusters of very unequal size (e.g. 10 points and 100 points)? How would you generalize the K-means algorithm to handle clusters of unequal size?

18. What happens if you run K means on data from two very elongated elliptical clusters as shown below? How would you generalize the K-means algorithm to handle elongated clusters?