8   Machine Learning and Bayesian Inference (sbh11)

You are faced with the following simple inference problem. Examples are pairs (x, y), where both x and y are real numbers. You suspect that there is a relationship $y = \theta x$ underlying the data, and that the y values are subject to additive Gaussian noise with mean 0 and variance σ^2. You wish to infer θ from examples (x_i, y_i) where $i = 1, \ldots, m$. You may consider x values to be fixed.

(a) Write down an expression for the density $p(Y|\theta; x)$. You may use the standard expression for the Gaussian density $N(\mu, \sigma^2)$

$$p(Z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Z - \mu)^2 \right).$$

[1 mark]

(b) Write down an expression for the likelihood of the parameter θ, given the m training examples described. State any assumptions you make.   [3 marks]

(c) Assuming that the parameter θ has a Gaussian prior with mean 0 and variance σ_1^2, write down an expression for the posterior density $p(\theta|\mathbf{y}; \mathbf{x})$ of θ. You may leave this expression unnormalized. State any assumptions you make.

[3 marks]

(d) Explain how the calculation in Part (c) might be extended to obtain the *evidence*, and how this might be used to estimate the values of σ and σ_1 from the training data. State any assumptions needed. You need not perform the actual calculation.   [3 marks]

(e) Find an expression for the *predictive density* $p(Y|x, \mathbf{x}, \mathbf{y})$.   You need not normalize the expression. You may use the identity

$$\int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}\left(\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{b} + c\right) \right) = (2\pi)^{p/2}|\mathbf{A}|^{-1/2} \exp\left(-\frac{1}{2}\left(c - \frac{\mathbf{b}^t \mathbf{A} \mathbf{b}}{4} \right) \right).$$

[7 marks]

(f) What further steps are necessary should you wish to extend your result from Part (e) to obtain actual predictions for Y along with a measure of certainty in the prediction?   [3 marks]