Machine Learning and Bayesian Inference (sbh11)

In designing a system to perform linear regression with noisy data, you feel that the noise is not modelled well by the usual normal density, and wish instead to use the Cauchy density

\[ p(x) = \frac{1}{\beta \pi} \left( \frac{\beta^2}{(x - \alpha)^2 + \beta^2} \right) \]

having parameters \( \alpha \) and \( \beta \).

(a) Denote the weights of your model by the vector \( \mathbf{w} \). Given a set \( \mathbf{s} \) of \( m \) examples, each consisting of a \( d \)-dimensional vector \( \mathbf{x} \) and corresponding label \( y \), find an expression for the likelihood \( p(y|\mathbf{w}) \) where \( y^T = (y_1, \ldots, y_m) \). State any assumptions you make. [6 marks]

(b) In addition to the unusual noise density, you have some knowledge of the problem at hand suggesting that some of the parameters in \( \mathbf{w} \) are likely to be close to known values, whereas the others have no such constraint. Suggest a suitable prior density \( p(\mathbf{w}) \) that could be used to model this. You should assume that the weights are independent for the purposes of designing a prior, and you may use the fact that the normal density is

\[ p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (x - \mu)^2 \right). \]

[5 marks]

(c) Using your answers to parts (a) and (b) derive a maximum a posteriori (MAP) learning algorithm for the problem. Should your algorithm require derivatives you may state them without working them out in full. [6 marks]

(d) Suggest a way in which any parameters other than \( \mathbf{w} \) might be either given an effective value or removed from consideration. Give a single advantage and a single disadvantage of the method you suggest. [3 marks]