Example Question for 3F3: Pattern Processing

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

1. Consider a linear classifier of the following form:

$$\begin{aligned} P(y=1|x,\theta,\epsilon) &= f_\epsilon(\theta x) \\ P(y=0|x,\theta,\epsilon) &= 1 - f_\epsilon(\theta x) \end{aligned}$$

where $y \in \{0,1\}$, $x \in \Re$, $0 \leq \epsilon \leq \frac{1}{2}$ and

$$f_\epsilon(z) = \epsilon + \frac{1-2\epsilon}{1+e^{-z}}$$

   (a) Sketch the function $f_\epsilon(z)$ for z between -3 and 3, showing ϵ on the sketch. [15%]

   (b) Write down the log likelihood as a function of the model parameters θ and ϵ for a data set $\mathcal{D} = \{(x_1, y_1) \ldots (x_N, y_N)\}$. [15%]

   (c) Calculate the derivatives of the log likelihood with respect to ϵ and θ for a data set $\mathcal{D} = \{(-1, 0), (+1, 1)\}$. [35 %]

   (d) Consider data in which the classes completely overlap: for example, the distribution of x for points in the $y = 1$ class is Uniform between -2 and 2, and likewise, the distribution of x for points in the $y = 0$ class is Uniform between -2 and 2. What values for ϵ and θ are appropriate in this case? [20 %]

   (e) Discuss the role of ϵ in dealing with data points with noisy labels. [15 %]