Homework 2 solutions

Problem 1

(a)

$$\frac{\partial \mathcal{L}}{\partial \pi} = 0 = \sum_{i=1}^{n} \frac{y_i}{\pi} - \frac{1-y_i}{1-\pi} \Rightarrow \pi = \frac{1}{n}\sum_{i=1}^{n} y_i$$

(b) Taking $\lambda_{1,d}$ as example

$$\frac{\partial \mathcal{L}}{\partial \lambda_{1,d}} = 0 = \frac{1}{\lambda_{1,d}} - 1 + \sum_{i=1}^{n}(\frac{y_i x_{i,d}}{\lambda_{1,d}} - y_i) \Rightarrow \lambda_{1,d} = \frac{1+\sum_{i=1}^{n} y_i x_{i,d}}{1+\sum_{i=1}^{n} y_i}$$

To cover all possible cases, can equivalently use indicators for $y \in \{0,1\}$. Results in

$$\lambda_{y,d} = \frac{1+\sum_{i=1}^{n} 1(y_i = y)x_{i,d}}{1+\sum_{i=1}^{n} 1(y_i = y)}$$

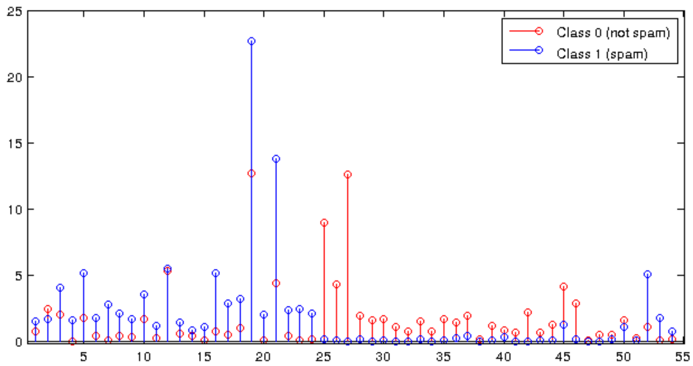
Problem 2

(a)

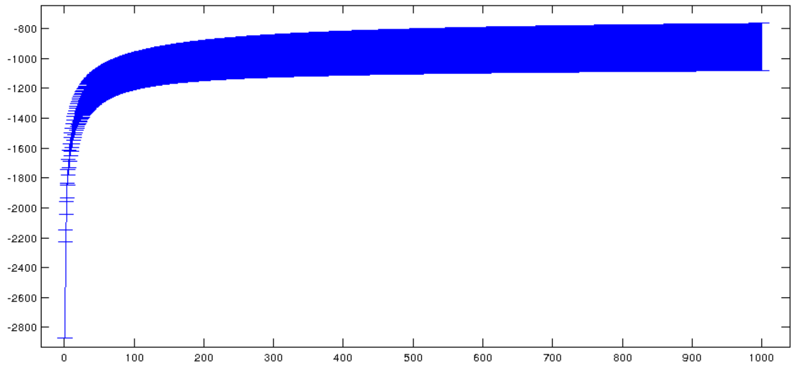
	pred = 0	pred = 1
truth = 0	2296±10	491±10
truth = 1	100±10	1713±10

prediction accuracy roughly = 0.8715.

(b) The stem plot is below. Dimension 16 is the word "free" and dimension 52 is "!". These two features make a spam email more likely.



(d) I ran it many times and plotted error bars below to give a sense of the range.



(e)

$$\mathcal{L}(w) \approx \mathcal{L}'(w) \equiv \mathcal{L}(w_t) + (w - w_t)^T \nabla \mathcal{L}(w_t) + \frac{1}{2}(w - w_t)^T \nabla^2 \mathcal{L}(w_t)(w - w_t)$$

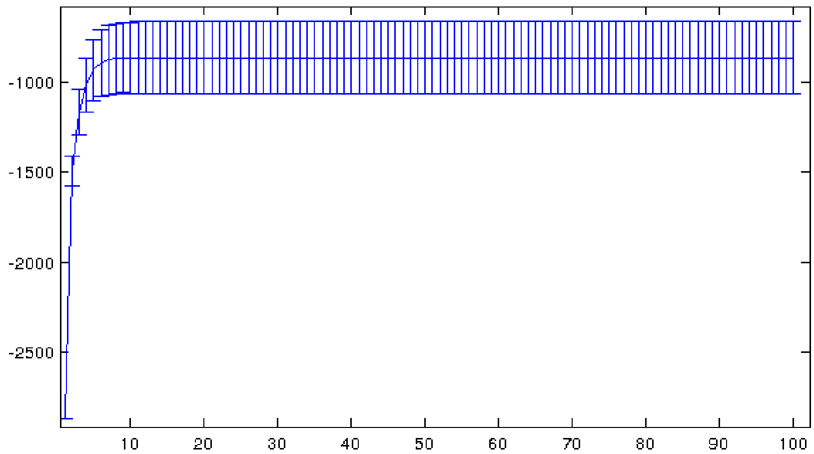
$$\nabla_w \mathcal{L}'(w) = 0 = \nabla \mathcal{L}(w_t) + \nabla^2 \mathcal{L}(w_t) w - \nabla^2 \mathcal{L}(w_t) w_t$$

$$w = w_t - (\nabla^2 \mathcal{L}(w_t))^{-1} \nabla \mathcal{L}(w_t) \Rightarrow w_{t+1}$$

$$\mathcal{L} = \sum_{i=1}^{n} \ln \sigma(y_i x_i^T w_t), \qquad \sigma(y_i x_i^T w_t) = \frac{1}{1 + \mathrm{e}^{-y_i x_i^T w_t}}$$

$$\nabla \mathcal{L}(w_t) = \sum_{i=1}^{n} (1 - \sigma(y_i x_i^T w_t)) y_i x_i$$

$$\nabla^2 \mathcal{L}(w_t) = -\sum_{i=1}^{n} \sigma(y_i x_i^T w_t)(1 - \sigma(y_i x_i^T w_t)) x_i x_i^T$$



(f)

	pred = 0	pred = 1
truth = 0	2645±12	142±12
truth = 1	214±12	1600±12