Bangwei Zhou (bz2280)

Professor Paisley

ELEN 4720 HW2

5 March 2020

1. **Solution to Problem 1**

   (a) $\hat{\pi} = argmax \sum_{i=1}^{n} lnp(y_i \mid \pi)$

   $= argmax \sum_{i=1}^{n} ln(\pi^{y_i}(1-\pi)^{1-y_i})$

   $= argmax \sum_{i=1}^{n} y_i ln\pi + (1-y_i)ln(1-\pi)$

   By FOC, $0 = \frac{\sum_{i=1}^{n} y_i}{\pi} - \frac{\sum_{i=1}^{n}(1-y_i)}{1-\pi}$

   $(1-\pi)\sum_{i=1}^{n} y_i - \pi \sum_{i=1}^{n}(1-y_i) = 0$

   $\pi n = \sum_{i=1}^{n} y_i$

   $\hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{n}$

   (b) We use an indicator and rewrite the equation as follows:

   $ln\lambda_{y,d} - \lambda_{y,d} + (\sum_{i=1}^{n}(x_i ln\lambda_{y,d} - \lambda_{y,d}))1\{y_i = y\}$

   By FOC, $0 = \frac{1}{\lambda_{y,d}} - 1 + (\frac{\sum_{i=1}^{n} x_i}{\lambda_{y,d}} - \sum_{i=1}^{n})1\{y_i = y\}$

   $1 - \lambda_{y,d} + \sum_{i=1}^{n} x_i 1\{y_i = y\} - \lambda_{y,d} \sum_{i=1}^{n} 1\{y_i = y\} = 0$

   $\hat{\lambda_{y,d}} = \frac{\sum_{i=1}^{n} x_i 1\{y_i=y\}+1}{\sum_{i=1}^{n} 1\{y_i=y\}+1}$

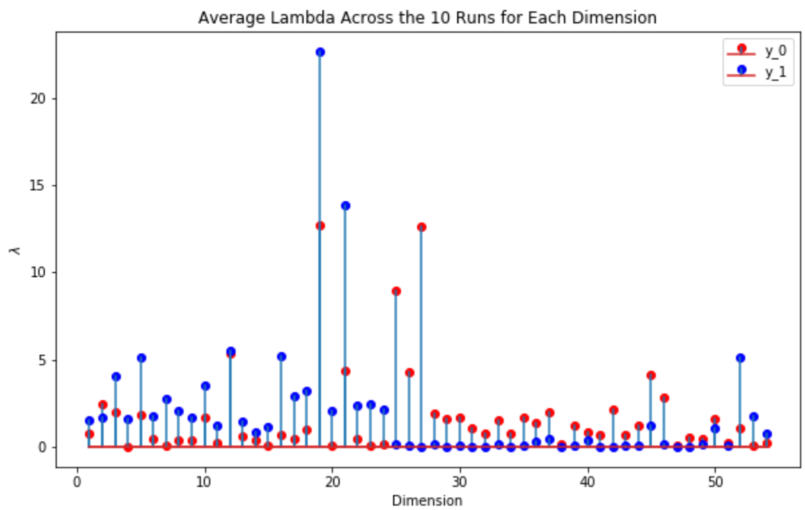
2. **Solution to Problem 2**

   (a)

	y_1	y_0
$pred_1$	1714	492
$pred_0$	99	2295

   prediction accuracy : 0.8715

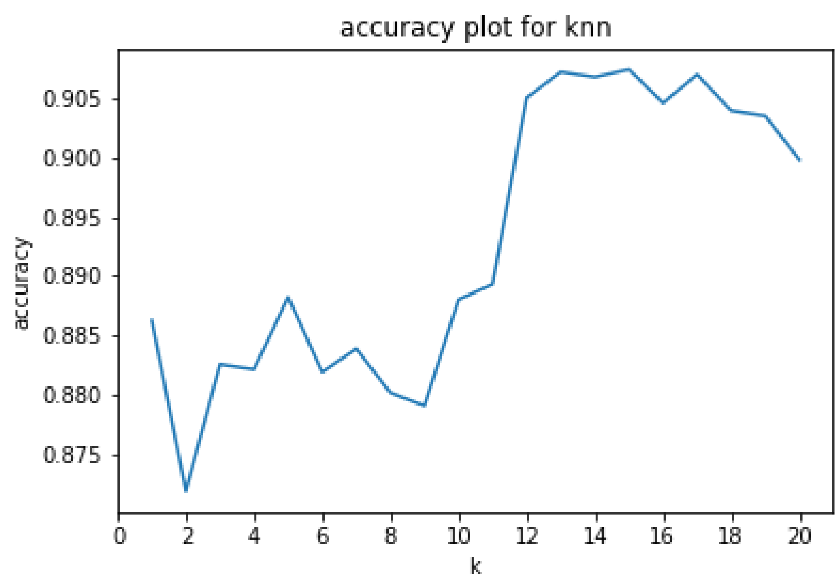
(b)



d_{16} is 'free' and d_{52} is '!'. They both have higher lambda for y when y = 1, than when y = 0, meaning 'free' and '!' appear more often in spam emails than non-span emails, as pois(lambda) models the frequency of these words. This seems to suggest that emails with 'free' and '!' are likely to be spam emails, which is quite intuitive and valid.

(c)

(d)



3. **Solution to Problem 3**

(a) Code in .py

(b)

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
5	1.96628	1.93314	1.92342	1.9222	1.92477	1.92921	1.93464	1.94059	1.94682	1.95321
7	1.92016	1.90488	1.90808	1.9159	1.92481	1.9337	1.94226	1.95038	1.9581	1.96544
9	1.89765	1.90252	1.91765	1.93252	1.9457	1.95724	1.96741	1.97649	1.98474	1.99234
11	1.89051	1.91498	1.93885	1.95794	1.97322	1.98577	1.99638	2.00561	2.01384	2.02135
13	1.89585	1.93559	1.9646	1.9855	2.00132	2.01388	2.02431	2.03331	2.04132	2.04864
15	1.90961	1.95955	1.99081	2.01192	2.02737	2.03947	2.04947	2.05811	2.06585	2.07298

where $b \in \{5, 7, 9, 11, 13, 15\}$ and $\sigma^2 \in \{.1, .2, .3, .4, .5, .6, .7, .8, .9, 1\}$

(c) The best value is 1.89 when b = 11, $\sigma^2 = 0.1$; the result is better than the first homework (around 2.2). The drawback might be the computation time. We are inverting nxn matrix here whereas in hw1 we are inverting dxd matrix. Thus, if n grows, computation might become expensive. Also feature selection is easier and more intuitive with ridge regression as compared to here.

(d)