COMS 4771 Fall 2016 Homework 2

Solutions by Daniel Hsu

Problem 1

Solution.

(a) $$\hat{\mu}_{y,j} \;:=\; \frac{\sum_{i=1}^n \mathbb{1}\{y_i = y \wedge x_{i,j} = 1\}}{\sum_{i=1}^n \mathbb{1}\{y_i = y\}} \;=\; \frac{\#\text{ messages with label } y \text{ and term } j}{\#\text{ messages with label } y}.$$

(b) Training error rate: 0.21626. Test error rate: 0.37602.

The weight vector $\hat{\boldsymbol{w}}_y = (w_{y,1}, w_{y,2}, \dots, w_{y,d})$ for class y is given by $\hat{w}_{y,j} = \ln(\hat{\mu}_{y,j}/(1-\hat{\mu}_{y,j}))$, and the threshold $\hat{t}_y$ for class y is given by $-\sum_{j=1}^d \ln(1-\hat{\mu}_{y,j}) - \ln(\hat{\pi}_y)$. Here, $\hat{\pi}_y$ and $\hat{\boldsymbol{\mu}}_y = (\hat{\mu}_{y,1}, \hat{\mu}_{y,2}, \dots, \hat{\mu}_{y,d})$ are, respectively, the class prior and class conditional distribution parameters for class y.

```
function params = hw2_train_bnb(X,Y)

% class "names"
params.classes = unique(Y);

K = length(params.classes);
params.w = zeros(size(X,2),K);
params.t = zeros(1,K);

for k = 1:K
  y = params.classes(k);

  % class conditional parameters
  mu = (1 + sum(X(Y == y,:))) / (2 + sum(Y == y));

  % weight vector and threshold
  params.w(:,k) = log(mu ./ (1-mu))';
  params.t(k) = -sum(log(1-mu)) - log(mean(Y == y));
end

function preds = hw2_test_bnb(params,test)
scores = bsxfun(@minus, test * params.w, params.t);
[~,I] = max(scores,[],2);
preds = params.classes(I);
```

(c) Training error rate: 0.05779. Test error rate: 0.13138.

(d) $\boldsymbol{\alpha} = \hat{\boldsymbol{w}}_{\text{political}} - \hat{\boldsymbol{w}}_{\text{religious}}$.

Most positive ("political"): firearms, occupied, israelis, serdar, argic, ohanus, appressian, sahak, melkonian, villages, cramer, armenia, cpr, sdpa, handgun, optilink, palestine, firearm, budget, arabs.

Most negative ("religious"): athos, atheism, atheists, clh, teachings, revelation, testament, livesey, atheist, solntze, wpd, scriptures, theology, believers, alink, ksand, benedikt, jesus, mozumder, prophet.

□

Problem 2

Solution. Let (X, Y) be a random example with the distribution P from the problem statement. For any $x \in \mathbb{R}$, the expected penalty of a classifier f conditional on $X = x$ is

$$c \cdot \mathbb{1}\{f(x) = 1\} \cdot P(Y = 0 \mid X = x) + \mathbb{1}\{f(x) = 0\} \cdot P(Y = 1 \mid X = x).$$

So the classifier f^* with smallest expected penalty is given by

$$f^*(x) \;:=\; \begin{cases} 1 & \text{if } P(Y = 1 \mid X = x) > c \cdot P(Y = 0 \mid X = x), \\ 0 & \text{otherwise}. \end{cases}$$

The condition in which $f^*(x) = 1$ is equivalent to the log-odds ratio at x being larger than c:

$$\frac{\frac{1}{3} \cdot \frac{1}{\sqrt{2\pi(1/4)}} e^{-\frac{(x-2)^2}{2(1/4)}}}{\frac{2}{3} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} \;>\; c.$$

This simplifies to

$$-\frac{3}{2}x^2 + 8x - (8 + \ln(c)) \;>\; 0.$$

By inspecting the quadratic expression on the left-hand side, we observe that the inequality has no solutions x when $c \geq e^{8/3} \approx 14.39$. When $c < e^{8/3}$, the solutions are given by the interval

$$I_c \;:=\; \left(\frac{8 - \sqrt{16 - 6\ln(c)}}{3}, \frac{8 + \sqrt{16 - 6\ln(c)}}{3} \right).$$

(a) $f^*(x) = 1$ if and only if $x \in I_c$ (defined above).

(b) $f^*(x) = 0$ for all $x \in \mathbb{R}$.

□

Problem 3

Solution.

(a) $\lambda_1 + \sigma^2, \lambda_2 + \sigma^2, \dots, \lambda_d + \sigma^2$

(a) $(\lambda_1 + \sigma^2)^{-2}, (\lambda_2 + \sigma^2)^{-2}, \dots, (\lambda_d + \sigma^2)^{-2}$

□