CS 189 Introduction to Machine Learning
Fall 2017

HW12

This homework is due **Tuesday, November 21 at 10pm.**

1 Getting Started

You may typeset your homework in latex or submit neatly handwritten and scanned solutions. Please make sure to start each question on a new page, as grading (with Gradescope) is much easier that way! Deliverables:

1. Submit a PDF of your writeup to assignment on Gradescope, "HW[n] Write-Up"
2. Submit all code needed to reproduce your results, "HW[n] Code".
3. Submit your test set evaluation results, "HW[n] Test Set".

After you've submitted your homework, be sure to watch out for the self-grade form.

(a) Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

None
Comment: this monster hw takes way more than 30 hours

(b) Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all sols are entirely in my words & that I have not looked at another stdt's sols. I have credited all external srcs in this writing

ProSlem # 2

a.)

$$J_\lambda(w) = \frac{1}{2}\|y - Aw\|_2^2 + \lambda\|w\|_1$$

$$= \frac{1}{2}(y-Aw)^T(y-Aw) + \lambda\sum_1^d |w_i|$$

$$= \underbrace{\frac{1}{2}y^Ty}_{g(y)} \underbrace{- \sum_1^d y_i A_i w_i - \frac{1}{2}w^TA^TAw - \lambda|w_i|}_{\sum_1^d f(A_i, y, w_i, \lambda)}$$

$$w^TA^TAw \rightarrow nw^Tw = \sum w_i^2 \times n$$

$$\frac{\partial J_\lambda(w)}{\partial w_i} = 0 \Rightarrow -y_iA_i + nw_i + \lambda\,\text{sign}(w_i) = 0$$

$$\Rightarrow w_i = \frac{y_iA_i - \lambda\,\text{sign}(w_i)}{n}$$

$\Rightarrow$ w* determined by the i-feature & the output regardless of other feature.

b.) $w_i^* > 0 \Rightarrow w_i^* = \dfrac{y_iA_i - \lambda}{n}$

c.) $w_i^* < 0 \Rightarrow w_i^* = \dfrac{y_iA_i + \lambda}{n}$

d.) $w_i^* = 0$ iff $y_i A_i = 0$ or $y_i A_i = \lambda$

e.)

$$J_\lambda(w) = \frac{1}{2}\|y - Aw\|_2^2 + \lambda\|w\|_2^2$$

$$\Rightarrow \frac{\partial J}{\partial w_i} = -y_i A_i + n w_i + 2\lambda w_i = 0$$

$$\Rightarrow w_i^* = \frac{y_i A_i}{n + 2\lambda}$$

λ cannot make w_i^* to be 0

$\Rightarrow$ L2 discourages w_i^* to be 0

$\rightarrow$ that's why L1 norm promotes sparsity

(or L2 norm discourage sparsity.

f.)

$\lambda = 0.000001$ is good

See code attached

Problem # 3

a.) See code attached

b.) See code & answer attached

c.) See code & answer attached

d.)

$$\Pr\{|z_i| \geq t\} \leq e^{-t^2/2\sigma^2}$$

$$t = 2\sigma\sqrt{\log d}$$

$$\Rightarrow \Pr\{|z_i| \geq 2\sigma\sqrt{\log d}\} \leq \underbrace{e^{-\frac{4\sigma^2\log d}{2\sigma^2}}}_{1/d}$$

$\forall i = 1..d$

$$\Rightarrow \max P\{|z_i| \geq 2\sigma\sqrt{\log d}\} \leq \frac{1}{d}$$

$$\Pr\{\max|z_i| \geq 2\sigma\sqrt{\log d}\} \leq \frac{1}{d}$$

e.) $\hat{w}_{top}$ returns the tops entries of $\hat{w}_{LS}$

but $E[\|\hat{w}_{LS} - w^*\|_2^2] = \sigma^2 \operatorname{trace}[(A^TA)^{-1}]$

$\Rightarrow$ $\hat{w}_{top}$ return the top s entries of $w^* + z'$

Problem # 6

a.) See code attached

b.) See code attached Basically:

age : take average

sex : Female = 0, male = 1

Discard ticket & cabin

Fare : take avg of same class

embarked : C = 1 0 0

Q = 0 1 0

S = 1 0 1

c.) See code attached

d.) See code attached

e.) See code attached

f.) See code attached

g.) See code attached

h.) skip

i.) skip

j) simple tree : titanic : Accuracy = 0.86

spam : 0.80

bagged : titanic : 0.86

spam : 0.84

random forest :

titanic : 0.669

spam : 0.709

k.) Submitted