50.007 Machine Learning 2015 Term 6

Date: 6th Nov 2015 Time: 2:30 PM Duration: 2 hours

Instructions to Candidates:

1. This paper consists of 11 questions with 5 printed pages (This title

page counts as the first page).

2. This is a closed book examination.

3. Cheet sheets are not allowed.

4. Answer all the questions.

5. Write your answers in the answer books provided.

6. Wish you success!

1

1. (2P) Multiple Choice: Ridge regression optimizes what objective func-

tion?

O mean average precision with squared l

2

ř

weights w

-norm }w}2 2

“

*D d“1*

*w2 d*

on

O mean squared error with squared l

2

ř

weights w

-norm }w}2 2

*D d“1*

w2 O mean squared “

*d*

on

error with l

1

-norm }w}

1

“

ř

*D d“1*

*|w*

*d*

| on weights w

2. (2P) Multiple Choice: The Support vector machine optimizes what ob-

jective function?

O mean zero-one error with squared l

2

ř

weights w

-norm }w}2 2

“

*D d“1*

*w2 d*

on

O mean squared error with squared l

2

ř

weights w

-norm }w}2 2

“

*D d“1*

*w2 d*

on

O mean hinge loss with squared l

2

ř

*w*

-norm }w}2 2

“

*D d“1*

*w2 d*

on weights

3. (1P) Multiple Choice: Neural Networks require feature mappings φ :

x ÞÑ φpxq in the lower layers to be designed by hand?

O Yes O No

4. (2P) Suppose you have two data samples x

1

P R3 with

*x*

*i*

̈

*,x*

2

̊ ̋

̨

‹ ‚,

then write down a matrix transformation A P R3ˆ3 such that the dis- tance }Ax

1

“

*x x x*

p1q i p2q i p3q i

*́ Ax*

2

} between sample x

1

and x

2

is equal to

*}Ax*

1

*b ́ Ax*

2

} “

*px*

p1q 1

*́ x*

p1q 2

q2 ` 4px

p2q 1

*́ x*

p2q 2

q2

5. (4P)

• given a set of ppfpx

1

*q,y*

1

*q,...,pfpx*

*N*

*q,y*

*N*

qq consisting predictions fpx

*i*

q on features x

*i*

and their corresponding ground truth labels y

*i*

, write down the formula for hinge loss averaged over the number of samples N

2

qq consisting predictions fpx

*i*

q on features x

*i*

and their corresponding ground truth labels y

*i*

, write down the formula for mean squared error (MSE) averaged over the number of samples N

6. (4P) compute some derivatives with respect to vector x: either com-

pute the gradient vector

∇fpxq “

̈

̊ ̊ ̊ ̋

̨

‹ ‹ ‹ ‚

or the linear mapping Dfpxqrhs which maps a direction vector h onto the directional derivative Dfpxqrhs of function f at point x. You can choose for each task separately whether you want to compute the gra- dient vector or the linear mapping. Hint: think where to apply product rule, and chain rule. Note that product rule is compatible with matrix multiplication, no matter whether the component is a matrix or a vector.

(a) fpxq “ 3x7 ́ 5x5 ` 7x2 ́ 3, x P R1 (b) fpxq “ xJAz

*Bf Bx Bf*

1

Bx . . .

2

*Bf Bx*

*d*

the r-th dimension ř

*d r“1*

xprq, of vector x P Rdˆ1,A x

P Rdˆd,z P Rdˆ1, where xprq is

(c) fpxq“pxJAxq4, x P Rdˆ1,A P Rdˆd (d) fpxq“}Ax}

2

*xJx, x P Rdˆ1,A P Rdˆd*

7. (1P) is this function convex, concave or none of both?

8. (1P) is this function convex, concave or none of both?

3

• given a set of ppfpx

1

*q,y*

1

*q,...,pfpx*

*N*

*q,y*

*N*

9. (2P) A task for thinking a bit: Why is the following measure no good objective function for measuring the error in a regression problem ? The error is computed between ground truth y

*i*

and prediction fpx

*i*

q as given by the function

*E “*

*N 1*

*Nÿ*

i“1

q3

Hint: you can imagine what can happen if this objective is used with a linear model: fpx

*i*

*pfpx*

*i*

q ́ y

*i*

*q “ wJx*

*i*

.

10. (3P) A more complicated task: Consider the the following distribution function ( a special case of a so-called gamma distribution). It is de- fined for positive real numbers x ą 0.

ppxq “ cx2 expp ́βxqβ3

Note x, c and β are real numbers. c ą 0,β ą 0. Your task: compute the maximum likelihood estimator for parameter β for given data x

1

*,...,x*

*n*

. We assume that the samples x

*i*

are indepen- dent. Hint: compute the maximum likelihood estimator as the maximum of the log-likelihood, that means after applying a logarithm.

11. (3P) Another somewhat more complicated task: which of the following

functions is a kernel and which is not ? Give an argument why.

4

Suppose that x

*i*

̃

*x*

p1q

̧

*x*

*i*

are two-dimensional vectors.

*kpx*

1

“

p2q i

ˆ

25 0 0 9

̇

*x*

2

*kpx*

1

*,x*

2

q “ xJ 1

*,x*

2

q “ xJ 1

ˆ 25 0 kpx

1

0

ˆ

2 ̇ ́9

*x*

2

1 0 2

̇

*x*

2

Hint: Recall: if kpx

1

*,x*

2

q “ xJ 1

*,x*

2

q is a kernel function, then there must exist a mapping φ and a Hilbert space such that

*kpx*

1

*,x*

2

q“xφpx

1

*q,φpx*

2

qy

where

*xv*

1

*,v*

2

y

is the inner product in the Hilbert space for two vectors v

1

and v

2

, and

*kpx*

1

q}2

is the norm of φpx

1

*,x*

1

q“}φpx

1

q in this Hilbert space. Check whether you can find an explicit mapping φ for the euclidean inner product, or whether you can find a contradictions to properties of a kernel.

**End of Paper**

5