IMPERIAL COLLEGE LONDON

TIMED REMOTE ASSESSMENTS 2020-2021

MEng Honours Degree in Electronic and Information Engineering Part IV
MEng Honours Degree in Mathematics and Computer Science Part IV
MEng Honours Degrees in Computing Part IV
MSc Advanced Computing
MSc Artificial Intelligence
MSc in Computing (Specialism)
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant assessments for the Associateship of the City and Guilds of London Institute

PAPER COMP70014=COMP97105=COMP97106

MACHINE LEARNING FOR IMAGING

Wednesday 17 March 2021, 10:00
Duration: 140 minutes
Includes 20 minutes for access and submission

Answer ALL TWO questions
Open book assessment

This time-limited remote assessment has been designed to be open book. You may use resources which have been identified by the examiner to complete the assessment and are included in the instructions for the examination. You must not use any additional resources when completing this assessment.

The use of the work of another student, past or present, constitutes plagiarism. Giving your work to another student to use constitutes an offence. Collusion is a form of plagiarism and will be treated in a similar manner. This is an individual assessment and thus should be completed solely by you. The College will investigate all instances where an examination or assessment offence is reported or suspected, using plagiarism software, vivas and other tools, and apply appropriate penalties to students. In all examinations we will analyse exam performance against previous performance and against data from previous years and use an evidence-based approach to maintain a fair and robust examination. As with all exams, the best strategy is to read the question carefully and answer as fully as possible, taking account of the time and number of marks available.

Paper contains 2 questions

1 a Multiple choice questions

i) Which of the following two linear models has higher model complexity?

* A: $h_\Theta(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \Theta_3 x_1 x_2$
* B: $h_\Theta(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2^4 + \Theta_3 x_1^2$

A) Model A

B) Model B

C) Both have the same complexity.

ii) A logistic regression model taking raw RGB pixel values of images of size 64×64 as input features has which of the following number of parameters:

A) 65

B) 4,096

C) 4,097

D) 12,289

iii) Given the following convolutional layers:

```
conv1 = Conv2d(in=1, out=1, kernel=3, stride=1, pad=0)
conv2 = Conv2d(in=1, out=1, kernel=5, stride=1, pad=0)
conv3 = Conv2d(in=1, out=1, kernel=2, stride=2, pad=0)
```

What is the size of the receptive field for the following forward pass?

```
conv1(conv3(conv1(conv3(conv2(x)))))
```

A) 16×16

B) 20×20

C) 26×26

D) 28×28

iv) Given the following neural network layers:

```
conv1 = Conv2d(in=1, out=2, kernel=5, stride=1, pad=0)
conv2 = Conv2d(in=2, out=4, kernel=3, stride=1, pad=0)
conv3 = Conv2d(in=4, out=8, kernel=3, stride=1, pad=1)
fc = Linear(in=???, out=128)
```

What is the value of the input parameter of the fully connected layer `fc` in the following forward pass that takes images of size 128×128 as inputs?

```
x = max_pool(conv1(x), kernel=2)
x = max_pool(conv2(x), kernel=2)
x = max_pool(conv3(x), kernel=2)
x = fc(linearize(x))
```

A) 120

B) 225

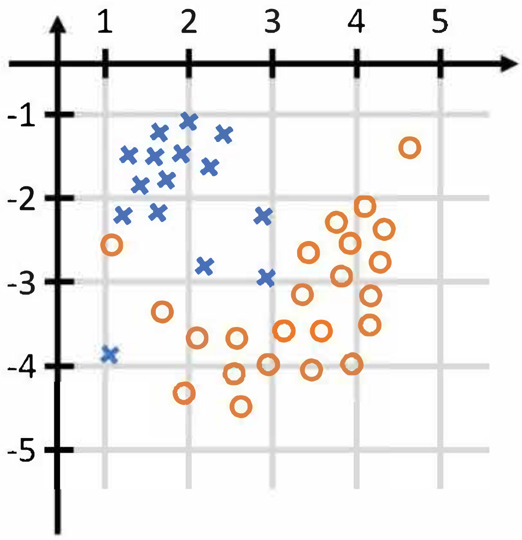
C) 1,800

D) 28,800

b Decision boundaries

i) Given the logistic regression model $h_\Theta(\mathbf{x}) = g(\Theta_0 + \Theta_1 x_1 + \Theta_2 x_2)$ and the following training data, draw a decision boundary that best separates the two classes. Determine the values for the parameters Θ that correspond to your decision boundary.

Hint: Try integer-valued crossing points on the coordinate axes.



ii) Assuming the circles in the above correspond to the positive class, and the crosses to the negative class, calculate the accuracy, precision, recall and specificity for your decision boundary.

c (Dis)similarity measures

Calculate the SSD, SAD, CC, and NMI for the following pair of images. Write down any intermediate values that you may calculate in the process.

Hint: Use a bin size of 1 for the joint intensity histogram.

2	1	2
2	4	3
5	4	1

1	2	1
1	-1	0
-2	-1	2

d Upsampling

Upsample the given 2×2 feature map by a factor of 2 using the given weight kernel and a nearest-neighbour-fill strategy. Fill in all cells marked with '?'.

2	4
-1	1

⇨

0	0	0	0	0	0
0	?	?	?	?	0
0	?	?	?	?	0
0	?	?	?	?	0
0	?	?	?	?	0
0	0	0	0	0	0

$*$

1	1	1
2	2	2
0	1	0

$=$

?	?	?	?
?	?	?	?
?	?	?	?
?	-3	?	?

e Forward pass

Given the following 1D input image and convolutional kernel:

3	1	-1	-3

$*$

-3	1	-3

calculate the output image for the following forward pass:

```
avg_pool(relu(conv1D(x, c, padding=1)), kernel=2)
```

The five parts carry, respectively, 20%, 20%, 30%, 15%, and 15% of the marks.

2 a Multiple choice questions

i) Which of the following statements regarding Haar features is correct? (Check all that apply)

- ☐ **Haar features can be computed efficiently via integral images**
- ☐ **Haar features are especially useful for characterising texture features**
- ☐ **The computational complexity of computing Haar features from the corresponding integral images is constant**
- ☐ **Haar features can approximate any filter which can be approximated via convolution**
- ☐ **None of the above**

ii) Which of the following is an advantage of end-to-end learning? (Check all that apply)

- ☐ **It is well suited for learning tasks with small dataset sizes**
- ☐ **It does not require hand crafted features**
- ☐ **It is efficient in terms of model parameters**
- ☐ **It learns models that are easily interpretable for humans**
- ☐ **None of the above**

iii) Which of the following can be used for regularisation in a neural network? (Check all that apply)

- ☐ L_1 **penalty on the weights**
- ☐ **Dropout**
- ☐ **Stochastic Gradient Descent with small batch sizes**
- ☐ **Stochastic Gradient Descent with large batch sizes**
- ☐ **Data augmentation**

iv) Which activation function is likely to lead to vanishing gradients? (Check all that apply)

- ☐ **Sigmoid**
- ☐ **tanh**
- ☐ **ReLU**
- ☐ **Parametric ReLU**
- ☐ **All of the above**

b Object detection

i) A particular version of YOLO uses 11 x 11 cells for object detection across images. Furthermore, it uses four object bounding box predictions per cell and it can detect 10 classes of objects. What is the size of the output that the CNN has to predict? Explain your answer.

ii) You now modify this particular version of YOLO to be able to detect 1000 object classes. How does the size of the output that the CNN has to predict change?

iii) Briefly explain what steps you could take to avoid the computational difficulties that may arise from increasing the number of classes to 1000.

c Super-resolution

Given a pair of ground truth image (below left), a super-resolution CNN produces a predicted output (below right).

8	8	3	7
1	4	5	6
2	8	6	2
6	7	0	9

Ground truth x

8	7	3	8
1	3	5	4
2	6	5	2
7	5	0	8

Prediction $\hat{x}$

i) Compute the value of the L1 loss as well as the L2 loss for the prediction.

ii) Assume a pre-trained CNN θ which consists of two different 2×2 filters f_1, f_2 shown below:

0	1
0	1

f_1

1	1
0	0

f_2

Assuming no padding and a stride of 2, compute the value of the perceptual loss for image pair above.

d Interpretability and Trustworthy ML

i) Assume a convolutional neural network which uses ReLU activation functions. Assume a feature map at layer l as shown below:

-3	3	9	8
2	-6	8	1
3	-2	-3	-1
0	5	1	3

Compute the result of the forward pass after the ReLU stage at layer l.

ii) In the corresponding backpropagation pass the gradient that is backpropagated from layer $l+1$ to layer l is shown below

2	8	4	-3
1	9	-5	3
2	1	-2	0
2	6	-1	4

Compute the values that are propagated in the backward pass to layer l for the following attribution methods:

a. Backprop
b. Deconvnet
c. Guided Backprop

iii) Given the image x and the gradient $\frac{\partial \mathcal{L}}{\partial x}$ shown below, compute the resulting adversarial example using the Fast Gradient Sign Method. Assume $\epsilon = 0.01$.

8	2	3	4
1	4	5	6
2	5	3	2
1	7	0	9

x

-7	2	1	5
1	-4	3	-9
9	-2	2	7
1	2	-1	3

$\frac{\partial \mathcal{L}}{x}$

iv) Explain why even though the magnitude of ϵ can be small, the Fast Gradient Sign Method can still be used successfully for adversarial attacks.

The four parts carry, respectively, 20%, 20%, 20%, and 40% of the marks.