IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2019

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 in Advanced Computing
MSc in Computing Science (Specialist)
for Internal Students of the Imperial College of Science, Technology and Medicine

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

PAPER C416

MACHINE LEARNING FOR IMAGING

Friday 22nd March 2019, 14:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 **Basics of machine learning**

a Logistic and linear regression

i) List the key elements in mathematical terms of supervised learning.

ii) State the general equation for linear regression for multiple variables.

iii) State the general equation for logistic regression for multiple variables.

b Optimisation

i) Sketch a pseudo-code implementation of gradient descent.

ii) Now extend this code to implement stochastic gradient descent.

iii) Draw a graph of the expected training and test error versus model complexity (e.g., depth of a decision tree). Describe briefly how the graph relates to bias and variance.

iv) Assuming that after training a model the performance evaluation indicates that there is high bias. Name three mechanisms for fixing high bias.

c Evaluation

Given the following 6×6 binary segmentations, with pixels marked with a '1' belonging to the foreground object, and pixels marked with '0' are background.

0	0	0	0	1	0
0	0	0	0	1	0
0	0	1	1	1	0
0	1	1	1	1	0
0	1	1	1	1	1
0	0	1	1	1	0

(a) Reference

0	0	0	0	0	0
0	1	0	0	0	0
0	1	1	0	1	0
0	1	1	1	0	0
0	1	1	0	1	1
0	0	0	0	0	0

(b) Prediction

i) Calculate the number of true and false positives, true and false negatives.

ii) Calculate the values for precision, recall, specificity, Dice similarity coefficient, and Hausdorff distance.

iii) Discuss briefly whether specificity is a useful metric for evaluating segmentation algorithms.

The three parts carry, respectively, 20%, 45%, and 35% of the marks.

2 MRFs, CNNs and FCNs

a Markov random fields

Given the following definition for the pairwise potentials of an MRF for image segmentation:

$$\psi_{ij}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j \\ \exp(-\frac{(y_i - y_j)^2}{2\sigma^2}) & \text{otherwise} \end{cases}$$

i) What is the general role of the pairwise potentials? What is the effect of this specific definition above?

ii) State clearly what each component in the above equation stands for.

b Convolutional neural networks

Given the following convolutional layers:

```
c1 = Conv2d(in=1, out=2, kernel=5, stride=1, pad=0)
c2 = Conv2d(in=2, out=4, kernel=3, stride=1, pad=1)
c3 = Conv2d(in=4, out=8, kernel=1, stride=2, pad=0)
c4 = Conv2d(in=8, out=4, kernel=5, stride=1, pad=1)
c5 = Conv2d(in=4, out=2, kernel=2, stride=2, pad=0)
c6 = Conv2d(in=2, out=2, kernel=3, stride=1, pad=0)
```

And the following forward pass: `x = c6(c5(c4(c3(c2(c1(x))))))`

i) What is the size of `x` after the forward pass for an input of size $1\times64\times64$?

ii) What is the size of `x` after the forward pass for an input of size $1\times128\times32$?

c Fully convolutional networks

Given a CNN designed for inputs of size $1\times112\times112$ with the six convolutional layers above plus two fully connected layers:

```
f1 = Linear(in=2*M, out=16)
f2 = Linear(in=16, out=10)
```

i) What is the missing value `M` in the layer `f1`?

ii) Convert the fully connected layers `f1` and `f2` into fully convolutional layers `c7` and `c8`. Specify the values of the parameters `in`, `out`, `kernel`, `stride` and `pad`.

d Upsampling

Upsample the given 2×2 feature map by a factor of 2 using the given weight kernel and a zero-fill (or “bed of nails”) strategy. Fill in all cells marked with ‘?’.

$$
\begin{array}{|c|c|}\hline -3 & 2 \\ \hline 4 & -1 \\ \hline \end{array}
\;\Rightarrow\;
\begin{array}{|c|c|c|c|c|c|}\hline 0&0&0&0&0&0\\ \hline 0&?&?&?&?&0\\ \hline 0&?&?&?&?&0\\ \hline 0&?&?&?&?&0\\ \hline 0&?&?&?&?&0\\ \hline 0&0&0&0&0&0\\ \hline \end{array}
\;*\;
\begin{array}{|c|c|c|}\hline 0.25&0.5&0.25\\ \hline 0.5&1&0.5\\ \hline 0.25&0.5&0.25\\ \hline \end{array}
\;=\;
\begin{array}{|c|c|c|c|}\hline ?&?&?&?\\ \hline ?&?&?&0.25\\ \hline ?&?&?&?\\ \hline ?&?&?&?\\ \hline \end{array}
$$

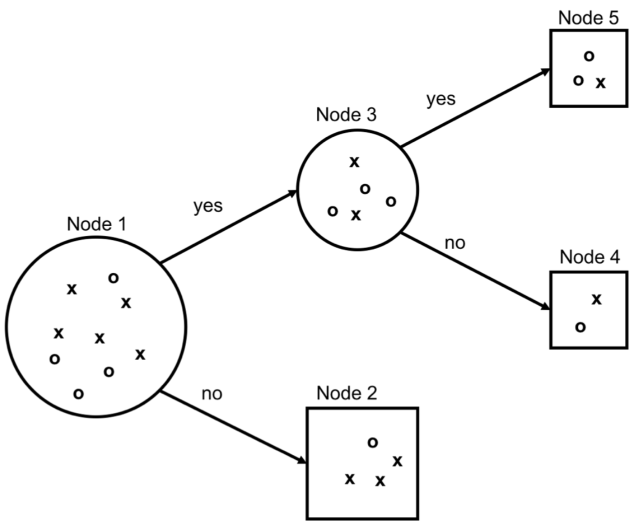
e Class imbalance

Name two strategies for handling class imbalance when training a model for image segmentation.

The five parts carry, respectively, 20%, 20%, 30%, 20%, and 10% of the marks.

3 Ensemble learning methods

a The figure below shows a binary decision tree during training with two classes marked by **o** and **x**.



Calculate the following:

i) The entropy at each of the 5 nodes.

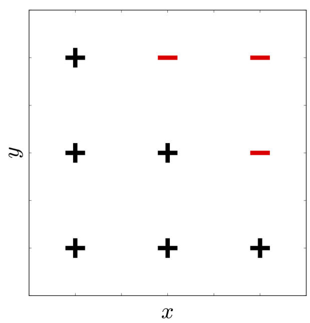
ii) The information gain at node 3.

iii) The overall information gain.

b Recall that Adaboost learns a classifier H using a weighted sum of weak learners h_t as follows

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

You are given the dataset below:



Furthermore, the question assumes decision trees as weak learners, which classify a point as $\{1, -1\}$ based on a sequence of *threshold* splits on its features (in this example x, y). To answer each of the questions below, replot the dataset and be sure to mark which regions are marked positive/negative. You can assume that ties are broken arbitrarily.

i) Assume that our weak learners are decision trees of depth 1 (i.e. decision stumps), which minimize the weighted training error. Draw the decision boundary learned by h_1.

ii) Circle the point(s) with the highest weights on the second iteration, and draw the decision boundary learned by h_2.

iii) Draw the decision boundary of $H = sign(\alpha_1 h_1 + \alpha_2 h_2)$. (Hint, you do not need to explicitly compute the α's).

iv) Now assume that our weak learners are decision trees of maximium depth 2, which minimize the weighted training error. Draw the decision boundary learned by h_1.

v) Circle the point(s) with the highest weights on the second iteration, and draw the decision boundary learned by h_2.

vi) Draw the decision boundary of $H = sign(\alpha_1 h_1 + \alpha_2 h_2)$. (Hint, you do not need to explicitly compute the α's).

c Does sequential ensemble learning reduce bias or variance? Motivate your answer.

The three parts carry, respectively, 35%, 50%, and 15% of the marks.

4 **Neural Networks**

a Consider a convolution layer. The input consists of 6 feature maps of size 20×20. The output consists of 8 feature maps, and the filters are of size 5×5. The convolution is done with a stride of 2 and zero padding, so the output feature maps are of size 10×10.

i) Determine the number of weights in this convolution layer.

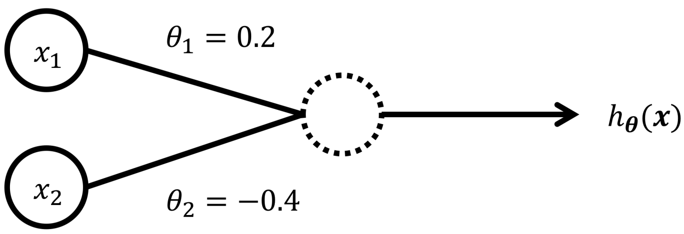
ii) Now suppose we made this a fully connected layer, but where the number of input and output units are kept the same as in the network described above. Determine the number of weights in this layer.

b The GoogLeNet (aka Inception V1) network uses inception blocks which make extensive use of 1×1 convolutions.

i) Sketch a typical inception block.

ii) Explain the purpose of the 1×1 convolutions in these inception blocks.

c Consider the following simple neural network:



Assume that the hidden unit in this network (dashed circle) uses a leaky RELU as activation function (i.e. $f_a(x) = \max(cx, x)$ with $c = 0.5$) and has no bias term (i.e. $\theta_0 = 0$). Furthermore, assume that the loss function used to train this network is $\mathcal{L} = \frac{1}{2}(h_\theta(\mathbf{x}) - y)^2$.

i) Show the computational graph for computing the loss function in a forward pass through the network.

ii) The network is trained with a single training example $\mathbf{x} = (2, 3)$ with label $y = 1$. Using the computational graph from part i) show the computations in the forward pass through the neural network.

iii) Given the training example above, show the computations for the back propagation with respect to θ_1 and θ_2 through the neural network.

d Regularisation can be used to avoid overfitting. Briefly describe two different strategies for regularisation when training neural networks. For each strategy describe how it avoids overfitting.

The four parts carry, respectively, 20%, 20%, 50%, and 10% of the marks.