SEMESTER 2 EXAMINATION 2018 - 2019

ADVANCED MACHINE LEARNING

DURATION 120 MINS (2 Hours)

This paper contains 4 questions

Answer all parts of the question in section A (30 marks)
and TWO questions from section B (35 marks each)

An outline marking scheme is shown in brackets to the right of each question.

This examination is worth 60%. The coursework was worth 40%.

University approved calculators MAY be used.

A foreign language dictionary is permitted ONLY IF it is a paper version of a direct 'Word to Word' translation dictionary AND it contains no notes, additions or annotations.

20 page examination paper.

Copyright 2019 © University of Southampton

Section A

Question A1.

(a) Explain the meaning of (1) the *prior distribution*, $\mathbb{P}(x)$, for a random variable, X, and (2) the **likelihood**, $\mathbb{P}(\mathcal{D}|x)$, of the data, $\mathcal{D}$, and write down (3) *Bayes' rule* for the posterior.

1

2

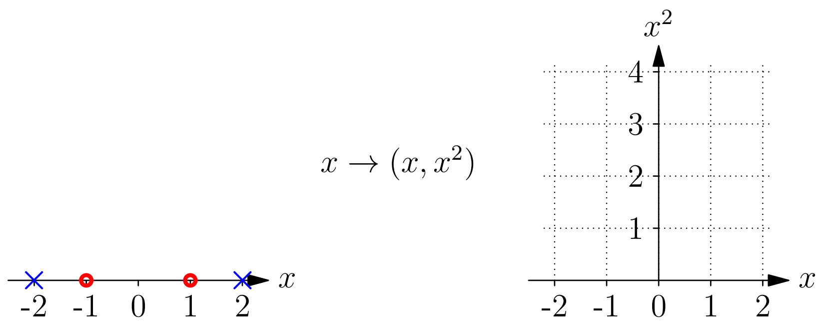
3

$\overline{6}$

[6 marks]

Copyright 2019 © University of Southampton

(b) For the one dimensional data points (crosses and circles) shown below, plot their position in an extended feature space created by the mapping $x \rightarrow (x, x^2)$. Draw the maximum margin dividing hyperplane in the extended feature space



[4 marks]

(c) Briefly describe the *Gradient Boosting* algorithm.

[5 marks]

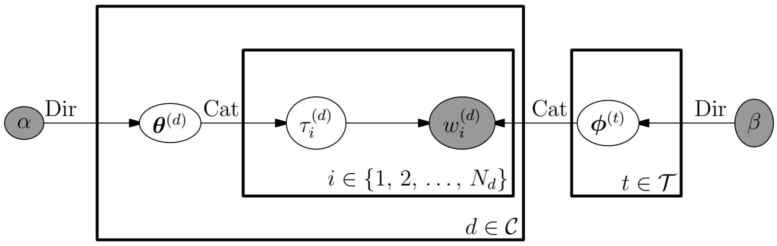
Copyright 2019 © University of Southampton

(d) Describe what you need to do practically to ensure that SVMs work well.

[5 marks]

$\overline{5}$

(e) The smoothed latent Dirichlet allocation topic model can be represented as a graphical model by the following plate diagram



where $\mathcal{C}$ is a set of documents and $\mathcal{T}$ is the set of topics. Sketch how documents of size N_d are generated by expanding the plate diagram

TURN OVER

Copyright 2019 © University of Southampton

to show the full word generation process.

$\overline{5}$

[5 marks]

Copyright 2019 © University of Southampton

(f) Show that the gamma distribution $\mathrm{Gam}(\mu|a,b) = b^a\,\mu^{a-1}\,\mathrm{e}^{-b\,\mu}/\Gamma(a)$ is a conjugate prior to the Poisson likelihood $\mathrm{Poi}(N|\mu) = \mu^N\,\mathrm{e}^{-\mu}/N!$ and derive the update equation for the parameters of the gamma distribution after observing N successes.

$\overline{5}$

[5 marks]

End of question A1

(a)	$\overline{6}$	(b)	$\overline{4}$	(c)	$\overline{5}$	(d)	$\overline{5}$	(e)	$\overline{5}$	(f)	$\overline{5}$	Total	$\overline{30}$

TURN OVER

Copyright 2019 © University of Southampton

Section B

Question B1.

(a) Explain for Gaussian Processes (GP) what is the prior, the likelihood and the posterior.

$\overline{}$ 5

[5 marks]

(b) Explain what the kernel function represents and how it could be measured empirically from many observations.

$\overline{}$ 5

[5 marks]

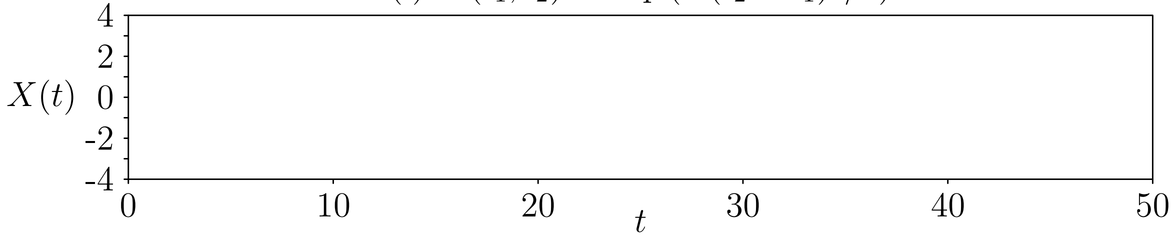
Copyright 2019 © University of Southampton

(c) Consider a 1-d Gaussian Process, $X(t)$, with a kernel of the form

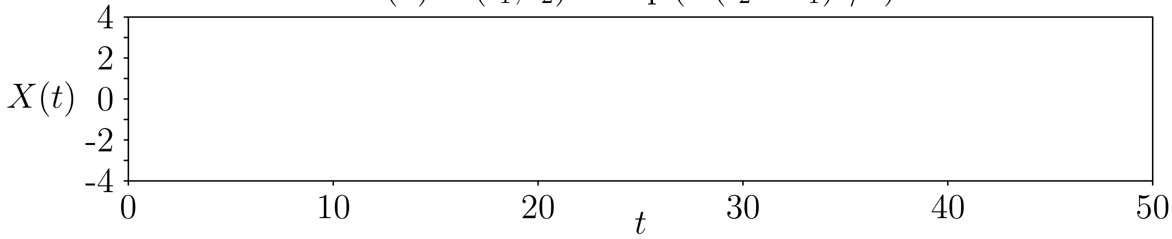
$$K(t_1, t_2) = \exp\left(-\frac{(t_2 - t_1)^2}{2\ell}\right).$$

Sketch three Gaussian Processes drawn from the prior with (i) $\ell = 1$ and (ii) $\ell = 2$ (we are not looking for accuracy, but rather the effect of changing ℓ).

(i) $K(t_1, t_2) = \exp\left(-(t_2 - t_1)^2/2\right)$



(ii) $K(t_1, t_2) = \exp\left(-(t_2 - t_1)^2/8\right)$



5

[5 marks]

Copyright 2019 © University of Southampton

(d) Explain the advantages and disadvantage of using the MAP solution rather than a full Bayesian solution.

$\overline{5}$

[5 marks]

(e) Explain why Monte Carlo techniques are often used to solve Bayesian inference problems.

5

[5 marks]

TURN OVER

Copyright 2019 © University of Southampton

(f) Briefly describe in words the use of the MCMC algorithm in Bayesian inference.

[5 marks]

$\overline{5}$

Copyright 2019 © University of Southampton

(g) When are probabilistic methods likely to give good results and what is the hurdle in using it?

$\overline{5}$

[5 marks]

End of question B1

(a) $\frac{}{5}$ (b) $\frac{}{5}$ (c) $\frac{}{5}$ (d) $\frac{}{5}$ (e) $\frac{}{5}$ (f) $\frac{}{5}$ (g) $\frac{}{5}$ Total $\frac{}{35}$

TURN OVER

Copyright 2019 © University of Southampton

Question B2.

(a) Show that the expected generalisation given by

$$\mathbb{E}_{\mathcal{D}}\big[E(\mathcal{D})\big] = \mathbb{E}_{\mathcal{D}}\left[\sum_{x \in \mathcal{X}} p(\boldsymbol{x}) \left(\hat{f}(\boldsymbol{x}|\mathcal{D}) - f(\boldsymbol{x})\right)^2\right]$$

can be written as the sum of a bias term, B, and variance term V where

$$B = \sum_{x \in \mathcal{X}} p(\boldsymbol{x}) \left(\hat{f}_m(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2, \quad V = \mathbb{E}_{\mathcal{D}}\left[\sum_{x \in \mathcal{X}} p(\boldsymbol{x}) \left(\hat{f}(\boldsymbol{x}|\mathcal{D}) - \hat{f}_m(\boldsymbol{x})\right)^2\right]$$

where $\hat{f}_m(\boldsymbol{x}) = \mathbb{E}_{\mathcal{D}}\left[\hat{f}(\boldsymbol{x}|\mathcal{D})\right]$ is the prediction made by averaging over all machines.

$\overline{15}$

[15 marks]

Copyright 2019 © University of Southampton

(b) Explain the *bias* and the *variance* terms in words.

1

2

$\overline{}$
5

[5 marks]

(c) What is the bias-variance dilemma.

$\overline{}$
5

[5 marks]

TURN OVER

Copyright 2019 © University of Southampton

(d) Explain how the *random forest* algorithm attempts to overcome the bias-variance dilemma.

$\overline{5}$

[5 marks]

(e) Explain how regularisation helps in the context of the bias-variance dilemma.

$\overline{5}$

[5 marks]

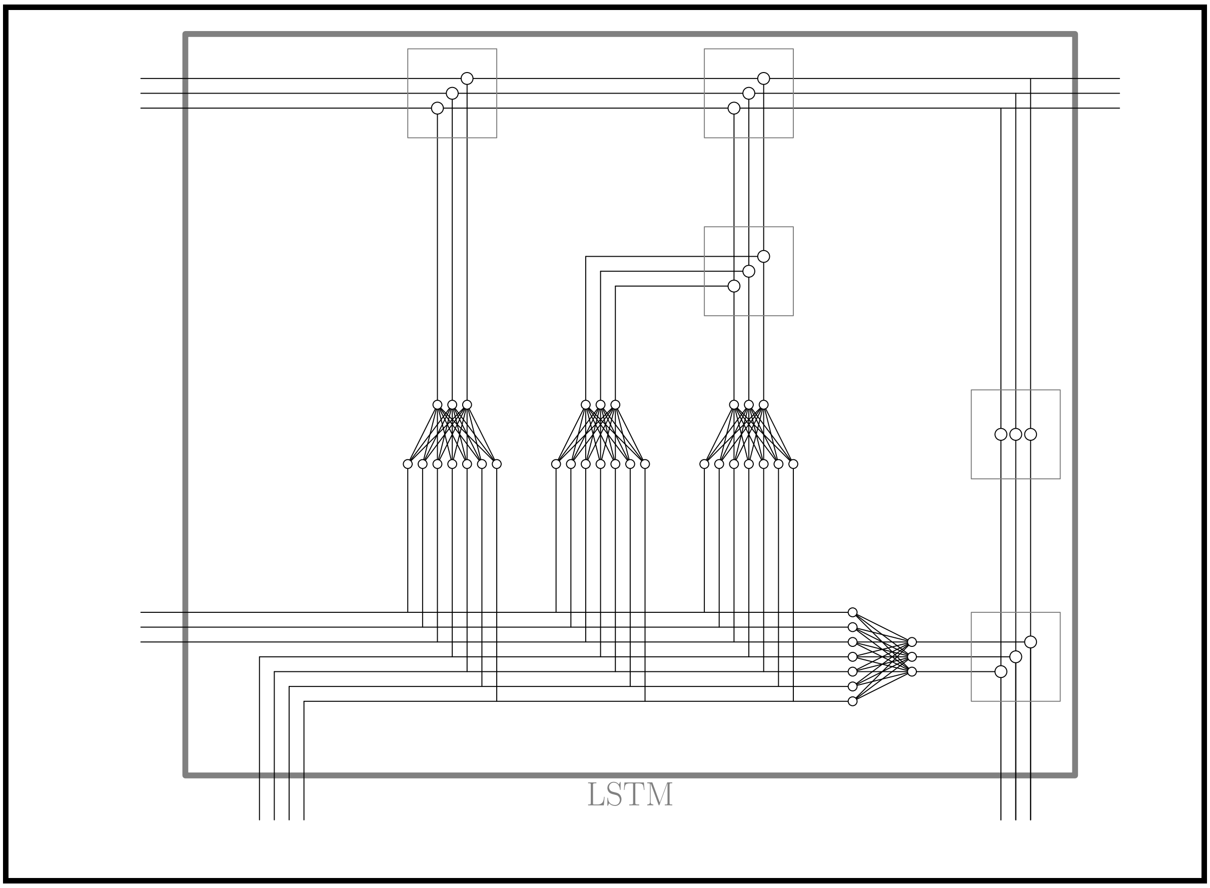
End of question B2

(a) $\overline{15}$ (b) $\overline{5}$ (c) $\overline{5}$ (d) $\overline{5}$ (e) $\overline{5}$ Total $\overline{35}$

Copyright 2019 © University of Southampton

Question B3.

(a) Add annotations to the figure below of an LSTM showing i) the memory $c(t-1)$ and $c(t)$, ii) the input $x(t)$, iii) the output $y(t-1)$ and $y(t)$, iv) the forget gate, v) the input/update gate vi) the output gate. In addition show whether the gates are multiplicative or additive and whether the nodes are sigmoidal (σ) or tanh function.



[15 marks]

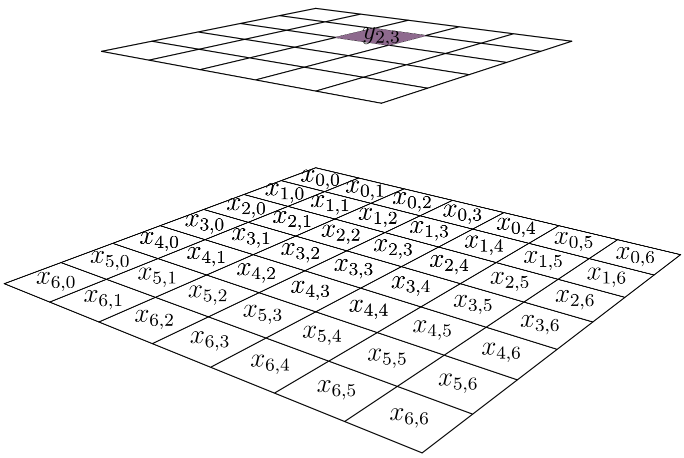
$\overline{15}$

Copyright 2019 © University of Southampton

(b) Explain what problem LSTM were designed to solve and how their architecture solves these problems.

[5 marks]

(c) In the figure shown below, the bottom layer describes an image and the top a convolution layer. Show the pixels that would contribute to the 3×3 convolution at $y_{2,3}$. Write down the value of $y_{2,3}$ in terms of the convolution filter f_{δ_x,δ_y} and the image pixel values $x_{i,j}$.



$y_{2,3} =$

[5 marks]

(d) Sketch the architecture of a residual network and explain what this architecture allows. Why are they seen to work where traditional CNNs fail?

5

[5 marks]

TURN OVER

Copyright 2019 © University of Southampton

(e) Describe what is meant by transfer learning in the context of CNNs.

$\overline{}$
5

[5 marks]

End of question B3

(a) $\frac{}{15}$ (b) $\frac{}{5}$ (c) $\frac{}{5}$ (d) $\frac{}{5}$ (e) $\frac{}{5}$ Total $\frac{}{35}$

END OF PAPER

Copyright 2019 © University of Southampton