A38624

Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Third Year Undergraduate

06-38965

LH Machine Learning

Main Summer Examinations 2025

Time allowed: 2 hours

[Answer all questions]

Note

Answer ALL questions. Note the appendix at the end of the exam paper.

Question 1 Core Knowledge

(a) Consider a logistic regression model with parameters $\mathbf{w} = (0.5, 0.1)^T$. What class would be predicted for an example described by inputs $\mathbf{x} = (1, 0.1)^T$? Show your calculations. **[12 marks]**

(b) Explain why one can claim that hard margin support vector machines use a strategy to reduce overfitting, and why this strategy may not always be sufficient.

[13 marks]

(c) The normal equation for regression is written as:

$$\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{w} - \mathbf{\Phi}^T\mathbf{t} = 0$$

Briefly describe the steps required to derive the normal equation. You do not need to reproduce the derivation itself. **[12 marks]**

(d) Explain the meaning of bias and variance in the context of a regression problem. What are the characteristics of a good learning model in terms of bias and variance? **[13 marks]**

Question 2 Support Vector Machines

Consider an example for which $yh(\mathbf{x}) = -3$, where $h(\mathbf{x}) = 0$ is the decision boundary learned by successfully solving the hard margin Support Vector Machine (SVM) optimisation problem with $\phi(\mathbf{x}) = \mathbf{x}$ using a given training set $\mathcal{T}$.

For each of the statements below, determine whether it is true or false, and justify your answer in detail.

(a) We know that this example is not in the training set $\mathcal{T}$. **[7 marks]**

(b) We know that this example is classified as being of class -1. **[6 marks]**

(c) We know that this example is misclassified by the SVM. **[6 marks]**

(d) We know that this example is on the margin. **[6 marks]**

Question 3 Learning Theory

We have a dataset of N=8 points $\{(x_i, t_i)\}_{i=1}^{8}$, where the independent variable is x_i and the dependent variable is t_i. Figure 1 below shows the results of fitting this dataset with a polynomial,

$$t_i = g(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{5} w_j x_i^j.$$

The line of best fit in each model is shown as a solid line. Each fit is generated by minimising a least squares loss function using gradient descent with or without regularisation. Each fit is the optimal result obtained by gradient descent using enough iterations.

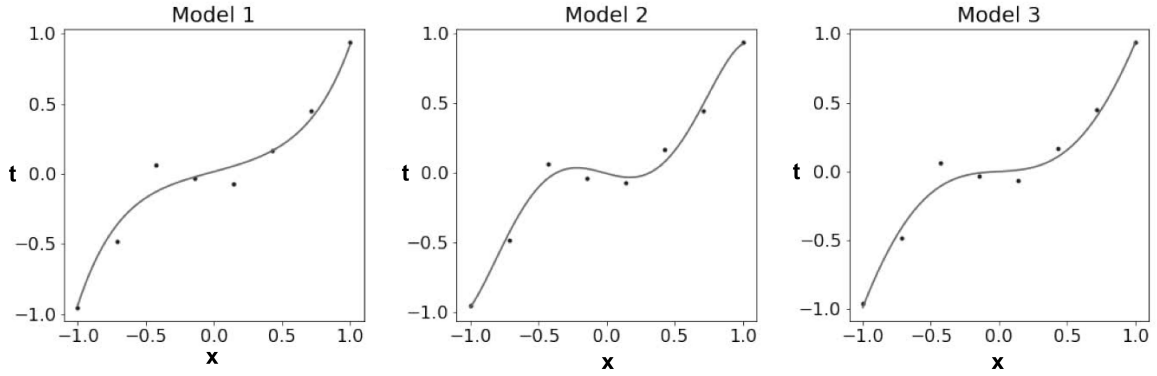


Figure 1: Fitting dataset with a polynomial.

The model parameters for each of the fits are listed in the table below:

	w_0	w_1	w_2	w_3	w_4	w_5
Model 1	0.015	0.265	0.018	0.378	-0.042	0.287
Model 2	-0.003	-0.261	0.170	2.389	-0.186	-1.184
Model 3	0.000	0.052	0.000	1.094	-0.027	-0.187

Table 1: Model parameters

Based on these results, for each of the statements below, determine whether it is true or false, and justify your answer in detail.

(a) From these fitting results, we can see that all three models have a good generalization ability. **[7 marks]**

(b) Model 1 is likely to be unregularised linear regression. **[6 marks]**

(c) Model 2 may or may not be unregularised linear regression. **[6 marks]**

(d) Model 3 is likely to be unregularised linear regression. **[6 marks]**

Appendix

Primal formulation of hard margin support vector machines

$$\underset{\mathbf{w},b}{\operatorname{argmin}} \left\{ \tfrac{1}{2} \|\mathbf{w}\|^2 \right\}$$

$$\text{subject to} \quad y^{(n)} h(\mathbf{x}^{(n)}) \geq 1, \quad \forall (\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T},$$

where $\mathbf{w}$ and b are the parameters to be learned, $\mathbf{x}^{(i)} \in R^d$ are the input variables of example i, d is the number of input variables, $y^{(i)} \in \{-1, 1\}$ is the output label of example i, $\mathcal{T}$ is the training set, $h(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b$, and $\phi(\mathbf{x})$ is a feature embedding. When $\phi(\mathbf{x}) = \mathbf{x}$, no embedding is being used.

Perpendicular distance between a point $\mathbf{x}^{(n)}$ and a hyperplane $h(\mathbf{x}) = 0$

$$dist(h, \mathbf{x}^{(n)}) = \frac{|h(\mathbf{x}^{(n)})|}{\|\mathbf{w}\|}$$

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.