

(a) You are given a dataset $\{(x_n, y_n)\}_{n=1}^N$ where $x_n \in \mathbb{R}^p$ is a p -dimensional vector of real-valued features associated with an individual n and y_n is a real-valued “output” variable for n . Show how you would set up a **linear regression** model by expressing y_n as a **linear** function of x_n . What would the learning algorithm return? [5 marks]

Assignment Project Exam Help

- (b) For a new dataset of input-output pairs $\{(x_n, y_n)\}_{n=1}^N$, with scalars $x_n, y_n \in \mathbb{R}$ you wish to express y_n as a cubic function of x_n , how would the corresponding design matrix \mathbf{A} be set up? For $N = 5$ data points describe the entries of the rows and columns of \mathbf{A} and introduce a loss function. [8 marks]

This image shows a single sheet of white paper with horizontal blue or grey ruling lines, typical of notebook paper. The paper is oriented vertically and has a small red pushpin pinned to its top edge, centered horizontally. The background behind the paper is dark and textured, suggesting it's resting on a desk or surface. The lighting is even, highlighting the texture of the paper and the lines.

[illegible]

- $$\{(x_1, y_1), (x_2, y_2)\} = \{(1.0, 0.5), (-1.0, -0.5)\}.$$

[12 marks]

Add WeChat powcoder

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

2

- (a) For a data set $\{x_n | n = 1, \dots, N\}$ of N column vectors x_n with feature values as components $x_{n,1}, x_{n,2}, \dots, x_{n,p}$, describe in detail how you would construct the covariance matrix of features. [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- (b) Using the singular value decomposition (SVD) of a $(N \times p)$ matrix $A = U\Sigma V^T$ describe how you would construct a rank k approximation of A , for $k < \min\{N, p\}$. [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- Assignment Project Exam Help
- <https://powcoder.com>
- Add WeChat powcoder

- (d) For classification problems where there is a training set of labelled data, explain why dimensionality reduction using PCA may give rise to poor classification outcomes. You may draw a figure to illustrate your arguments. [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

3

(a) Describe the k -means clustering algorithm.

[7 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- (b) In Fisher's Linear Discriminant Analysis (LDA), training data – \mathbf{x}_A^i and \mathbf{x}_B^j from 2 classes A and B – is projected along some vector \mathbf{w} so that $\mathbf{x}_A^i, \mathbf{x}_B^j$ are mapped into

$$y_A^i := \mathbf{w}^T \mathbf{x}_A^i, i = 1, \dots, N_A, \text{ and } y_B^j := \mathbf{w}^T \mathbf{x}_B^j, j = 1, \dots, N_B$$

respectively. Means and covariances are computed from the training data. These empirical vector means are denoted \mathbf{m}_A and \mathbf{m}_B and the variance-covariance matrices are called S_A and S_B respectively. Let μ_A and μ_B be the (scalar) means for the values $\{y_A^i\}$ and $\{y_B^j\}$, and σ_A, σ_B the corresponding variances, and will depend on the choice of vector \mathbf{w} . What is the quantity that LDA seeks to maximise in order to achieve classification accuracy? Explain why that is a good choice. Discuss whether the quantity is invariant under scaling. [5 marks]

Assignment Project Exam Help

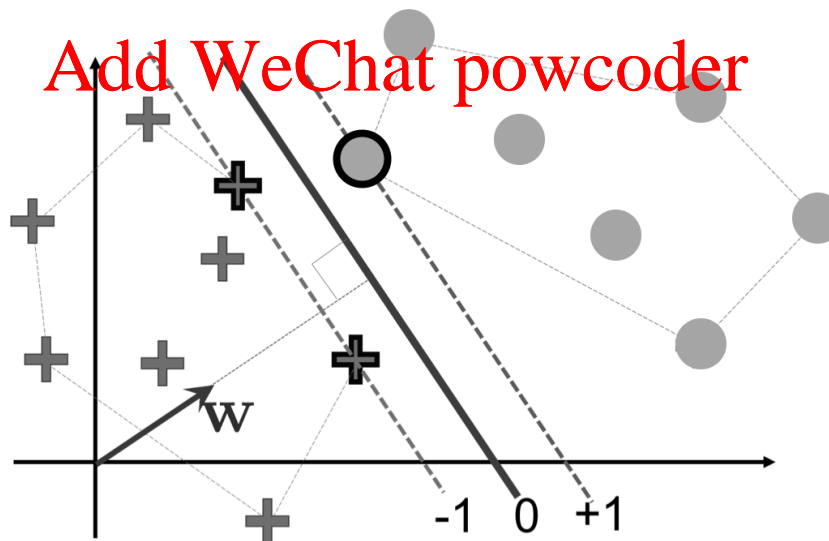
<https://powcoder.com>

Add WeChat powcoder

- (c) A linear hyperplane is described by the equation $y = w \cdot x + b$. The decision boundary in the figure is the line (representing a hyperplane) for which $y = 0$ (labelled 0) and is perpendicular to w . The two parallel hyperplanes that go through the support vectors (points with thickened edges) are indicated by the values $y = \pm 1$. Explain why a large margin is necessary for robust classification. From the geometry of the figure show that the size of the margin along the direction of w is $\frac{2}{\|w\|}$. You may take x_+ and x_- to be support vectors.

<https://powcoder.com>

Add WeChat powcoder



[7 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- $$\min_{\mathbf{w}, b} \max_{\boldsymbol{\alpha}} \mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{n=1}^N \alpha_n (y_n(\mathbf{w}^T \mathbf{x}_n + b) - 1), \quad \alpha_n \geq 0.$$

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

(a) For a probability distribution p over an event space \mathcal{X} $p := \{p_i | i \in \mathcal{X}\}$, explain why the entropy $H(p)$

is often interpreted as the average degree of ‘surprise’. [3 marks]

Add WeChat powcoder

- $$H(p, q) = \sum_i p_i \log\left(\frac{1}{q_i}\right) \text{ and } KL(p||q) = H(p, q) - H(p) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right).$$

Provide some intuition for what each of the two metrics over pairs of probability distributions captures. You may treat the distribution p as the one representing ‘ground truth.’ [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- (c) You are given a sequence $\mathbf{x} := \{x^{(1)}, \dots, x^{(N)}\}$ of heads ($x^{(i)} = H$) and tails ($x^{(i)} = T$) which are the outcomes of N tosses of a (potentially biased) coin. All possible outcomes of N coin tosses would constitute the event space \mathcal{X} . A binomial distribution $B(N, \theta)$ sets the probability of occurrence of n_1 events of type 1, and $n_2 = N - n_1$ of type 2 as

$$P(n_1, n_2 | N, \theta) = \frac{N!}{n_1! n_2!} \theta^{n_1} (1 - \theta)^{n_2}.$$

Describe how you would fit the data \mathcal{X} to a binomial distribution using maximum likelihood estimation and find the result to be the same as the empirical distribution $\tilde{p} = (\tilde{p}_H, \tilde{p}_T)$:

$$\theta^* = \frac{n_H}{N} =: \tilde{p}_H$$

[10 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- (d) A 2-dimensional Gaussian distribution is defined by the probability density function of $\mathbf{X} = (X_1, X_2)^T$ parameterised by its mean $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and variance-covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

$$p(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})\right).$$

Draw 3 contour plots of equiprobable values of (X_1, X_2) for the cases (a) $\rho = 0$, (b) $\rho < 0$ and (c) $\rho > 0$ providing reasons for doing so. [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

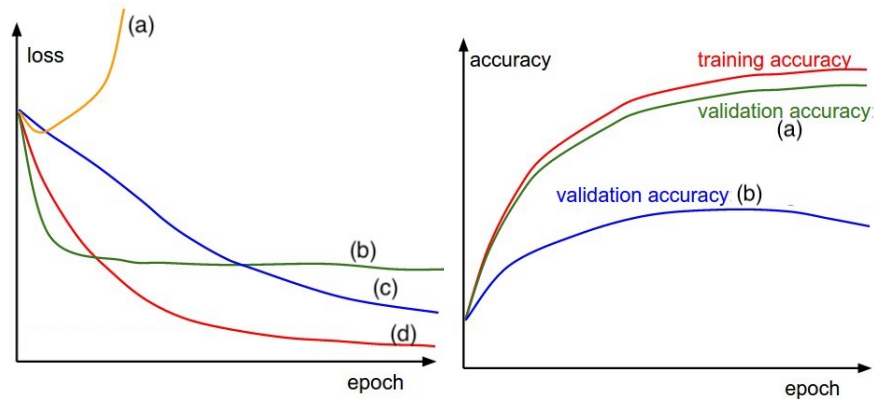
Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

5 Suppose we are training a neural network to do classification on the MNIST dataset. We run various experiments and plot the loss and accuracy over epochs and observe the plots shown in the figures below.



- (a) In order to generate the loss vs. epochs plot above, the learning rate is modified four times resulting in the curves marked (a), (b), (c) and (d). The trends behave quite differently as a result of modifying the learning rate. You need to decide which learning rate to use. Which case would you choose (a), (b), (c) or (d) and explain why. [6 marks]

<https://powcoder.com>

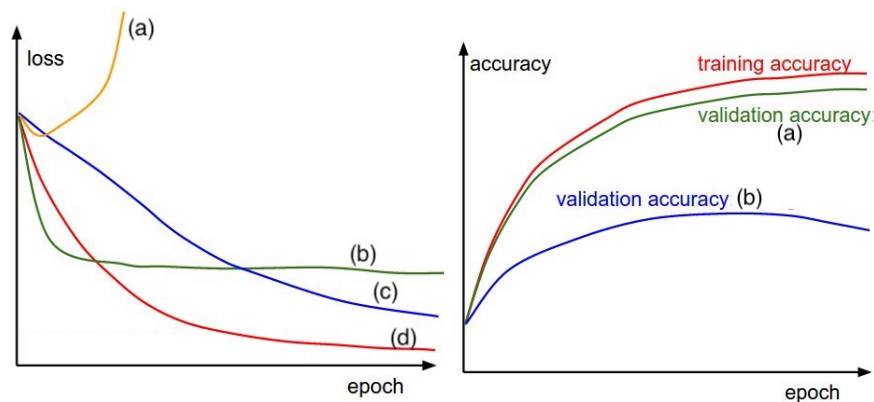
Add WeChat powcoder

- (b) Looking at the loss curve marked (a) in the loss plot, how can you modify your network to solve the issue? [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



- (c) Now consider the accuracy of your model on training versus validation sets over epochs and assume you observe the results shown in the above figure. Comparing the curves labeled (a) and (b) which are both plotting the validation accuracy, which of these cases illustrates overfitting? Can you explain why? [7 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- (d) How can you prevent your neural network model from overfitting? [6 marks]

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

END OF PAPER