Deep Learning/Deep Learning for Vision

Vineeth N Balasubramanian
Indian Institute of Technology, Hyderabad

ASSIGNMENT 0

This is a preparatory assignment for the course on "Deep Learning"/"Deep Learning for Vision". Below are questions on basics/pre-requisites for this course. If you can answer these, you are ready for the course!

1 Functions and Derivatives

1.1 Simple derivatives

Question: Find the derivative of $(\sin x + e^{2x} + \sqrt{x})$.

1.2 Activation function

Question: Activation functions are used by neural networks to learn non-linear decision boundaries. Popular activation functions used in neural networks are sigmoid, tanh and ReLU. Find the derivative of $\tanh(x)$. Recall that $\tanh(x)$ is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

1.3 Chain rule

Question: Apply chain rule to find the derivative of $\sin \frac{\sqrt{e^x+a}}{2}$. where a is constant.

2 Probability and Statistics

2.1 Bayes Theorem

Question: Using Bayes theorem, answer the following question. A laboratory blood test is 95% effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1% of healthy persons tested (i.e., if a healthy person is tested, then with probability 0.01, the test result will imply that he or she has the disease). If 0.5% of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive?

2.2 Conditional Dependence and Conditional Independence

Recall that two random variables X, Y are independent if $P(X,Y) = P(X)P(Y)$ or $P(X|Y) = P(X)$. Similarly we can define conditional independence as follows. Two random variables X, Y are conditionally independent given another random variable Z with $P(Z) > 0$, if $P(X,Y|Z) = P(X|Z)P(Y|Z)$ or $P(X|Y,Z) = P(X|Z)$. Two random variables may be unconditionally dependent but it is possible that they are conditionally independent and vice versa. Now, answer the following question.

Question: A box contain two coins. First coin is a fair coin ($P(H) = 0.5$). Second coin is a biased coin with $P(H) = 1$. Now, we choose a coin at random and toss it twice. Let us define the following events.

- A = First toss results in head
- B = Second toss results in head
- C = Fair coin is selected

Find $P(A), P(B), P(A \cap B), P(A|C), P(B|C), P(A \cap B|C)$ and identify the relationship between A, B, C.

2.3 Entropy, Cross Entropy and KL divergence

Question: Write the expression of KL divergence in terms of entropy and cross-entropy. Comment on the 'symmetry' of KL divergence (i.e. is $KL(P,Q) = KL(Q,P)$?).

Question: Find the KL divergence between two probability distributions P, Q where P, Q are as follows:

X	x_1	x_2	x_3
$P(X = x_i)$	0	0	1
$Q(X = x_i)$	0.25	0.5	0.25

2.4 Mutual Information

Question: Mutual Information of two random variables X, Y is defined as follows:

$$I(X;Y) = \sum_{x \in \mathbb{X}, y \in \mathbb{Y}} P_{(X,Y)}(x,y) \log \frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}$$

Prove that $I(X;Y) = H(Y) - H(Y|X)$, where $H(Y)$ is entropy of random variable Y and $H(Y|X)$ is conditional entropy of Y given X.

3 Linear Algebra and Matrix Operations

3.1 Norms of a vector

Question: Find L_1, L_2 and L_∞ norms of the vector $\mathbf{v} = \begin{pmatrix} 0.5 \\ -3 \\ -1 \\ 2 \end{pmatrix}$

3.2 Linear Independence, Rank

Question: Which of the following set of vectors are *linearly independent*?

$$\text{Set 1} = \left\{ \begin{pmatrix}4\\0\\0\end{pmatrix}, \begin{pmatrix}0\\4\\0\end{pmatrix}, \begin{pmatrix}0\\0\\4\end{pmatrix} \right\}$$

$$\text{Set 2} = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right\}$$

$$\text{Set 3} = \left\{ \begin{pmatrix}2\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}4\\3\\2\end{pmatrix} \right\}$$

Question: What is the rank of the following matrices?

$$M_1 = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}, M_2 = \begin{pmatrix}1 & 0 & 3\\0 & 1 & 4\end{pmatrix}, M_3 = \begin{pmatrix}1 & 2\\2 & 4\\3 & 6\end{pmatrix}$$

3.3 Eigenvalues and Eigenvectors

Question: Find the eigenvalues and eigenvectors of the matrix: $M = \begin{pmatrix}-2 & 1\\12 & -3\end{pmatrix}$

Question: If x is a vector, then prove that $x^T x$ is the eigenvalue of the matrix xx^T corresponding to the eigenvector x.

Question: Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of the matrix A corresponding to the eigenvectors $x_1, x_2, \ldots, x_n$. What are the eigenvalues of the matrix $A^2 + I$, where I is the identity matrix?

3.4 Singular Values and Singular Vectors

Question: Given a real matrix A of size $m \times n (m \geq n)$, what are its singular values and singular vectors, and how are they related to eigenvalues and eigenvectors of $A^T A$ and AA^T?

4 Numerical Methods

4.1 Taylor Series

Question: Write down the Taylor series expansion of a function $f(x)$ where $x \in \mathbb{R}$ and $f(x) \in \mathbb{R}$, and find the value of $e^{1.01}$ using a second-order approximation of its Taylor series.

Question: Write down the Taylor series of a function $f(\mathbf{x})$ where $\mathbf{x} \in \mathbb{R}^n$, $f(\mathbf{x}) \in \mathbb{R}$.

5 Basics of Machine Learning

5.1 L_1, L_2 regularization

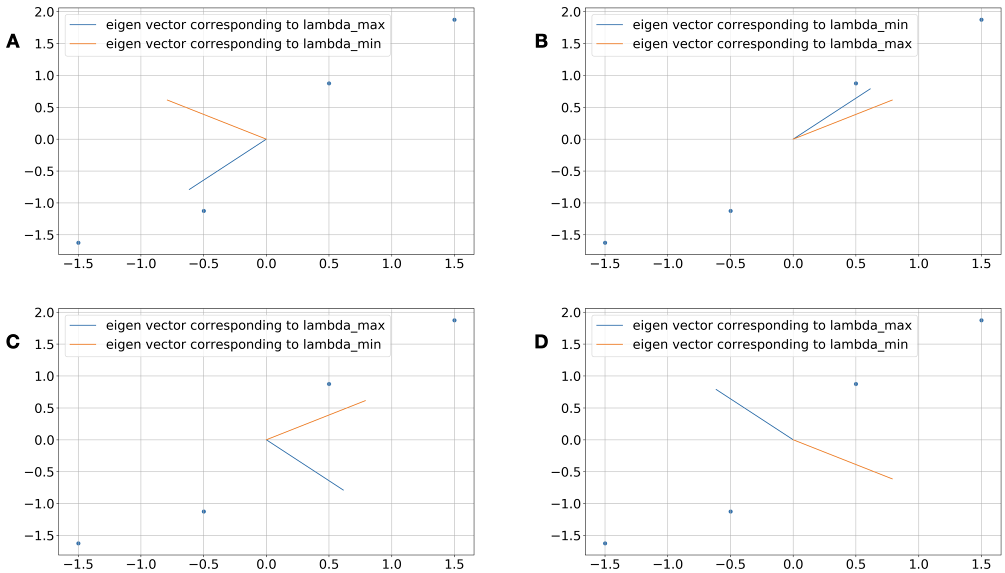
Question: Why does the L_1 regularizer lead to sparse models and L_2 regularizer does not?

5.2 Mean Squared Error and Least Squares Regression

Question: State and find the derivative of Mean Squared Error for Least Squares Regression.

5.3 Principal Component Analysis

Question: Suppose you are given a set of data points $P = \{(2,4),(1,3),(0,1),(-1,0.5)\}$ and you need to write a program to find the direction of maximum variance in the dataset. Which of the following plots is likely to be your output? In the plots, lamba_max and lambda_min represents maximum and minimum eigenvalues of covariance matrix of dataset (Zoom into the plots if required).



5.4 Logistic Regression

Question: Consider a binary classification problem. You have m data points in the form of (x_i, y_i) where x_i is i^{th} input point and $y_i \in \{0,1\}$ is the class label and you are using the sigmoid function to predict the probability of the class based on weights w and input x. Write down the logistic loss based on the given setting.

5.5 Kernels & SVM

Question: Prove that $K(x,y) = x^T y$ is a valid kernel.

5.6 Adaboost

Question: In the AdaBoost algorithm, if a decision stump (a decision tree of depth 1) misclassifies 1 out of 8 data points in one of the steps, what is the final importance of that stump in decision making?

5.7 Gaussian Mixture Models

Question: Which of the following statements about Gaussian Mixture Models is true?

1. Number of clusters $'k'$ in GMM is a learnable parameter.
2. GMM is a soft clustering technique.
3. Cluster assignment is done based on hard thresholding. That is, if the probability of a point belongs to a cluster is more than a specified threshold, that point is assigned to that cluster only.
4. GMM uses Expectation Maximization(EM) algorithm which makes GMM a Discriminative model.