Consider a two class classification problem in 2 dimensions. We know that both the classes can be modelled as multivariate Gaussians. We have 1000 samples each from both the classes (i.e., N=2000).

Bayesian Optimal Classifier gives 90% as the optimal accuracy.

We use a linear SVM.

(A) number of Support Vectors will be closer to 0.9 N.

(B) number of Support Vectors will be closer to 0.9 d.

(C) number of Support Vectors will be closer to 0.1 N.

(D) number of Support Vectors will be closer to 0.1 d.

(E) Bayesian optimal rate has no influence on the number of Support Vectors.

E? (intuitive)

ESTION IMAGE

"Since for a K class problem, DDAG uses ${}^{K}C_2$ classifiers, the final decision can be ambigous". (Write Tue or False)

Quiz 3, Question 2

Question Image

Consider a single layer perceptron with two input and one output. The weights from from first and second inputs are w_1 and w_2 respectively. Also assume a -1, +1 logic. Let w_0 be the weights associated with bias +1.

The activation at the output is:

$$\phi(x) = +1 \text{ if } x \geq 0 \text{ and } -1 \text{ else}$$

If $w_0 = 0, w_1 = 1, w_2 = 1$, then this perceptron is equivalent to:

(fill from the gates like: AND, OR, ExOR, NAND, NOR)

Question Description

Quiz 3, Question 4

Question Image

Consider the following 10 samples used for training a Kernel SVM with $\kappa(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T\mathbf{q})^2$. Labels are also given.

$$([-1,-1]^T, +1), ([1,1]^T, +1)([+3,+4]^T, +1), ([0,0]^T, -1)([10,10]^T, -1)$$

$$([0,1]^T, +1), ([-10,-10]^T, -1)([1,0]^T, -1), ([-2.5,-3.5]^T, +1), ([4.5,6.5]^T, -1)$$

corresponding α are:

$$0, 1, 0, 1, 0, 2, 0, 1, 0, 0$$

(α *values are scaled/adjusted to make the numerical computation simpler!*)
Assume $b = 0$.

Consider at the test time, we have a sample $[2,2]^T$ Is this sample in positive class or negative class?

Question Description

Quiz 3, Question 6

Question Image

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

The optimization problem that **Logistic Regression** solves is *convex*.

Consider two quadratic kernels: $\kappa_1(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T\mathbf{q} + 1)^2$ and $\kappa_2(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T\mathbf{q})^2$.

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

$\kappa_1(\cdot, \cdot)$ is a *valid kernel*; and $\kappa_2(\cdot, \cdot)$ is a *invalid kernel*;

Question Description

Quiz 3, Question 1

Question Image

Consider an MLP with two inputs, three hidden neurons and one output neurons. Hidden neurons and output neurons have sigmoid activation. There is no bias. Output neuron has a MSE loss.

Consider a sample $([5,5]^T, 0.7)$ i.e., $x = [5,5]^T$ and $y = 0.7$. We would like to update all the weights based on the gradient of the loss $(\mathcal{L})$. Assume that $w_{ij}^{[k]}$ connects ith neuron of layer k with jth neuron of layer k+1. Thus weights between input and hidden layer are $w_{11}^{[1]}, w_{21}^{[1]}, w_{12}^{[1]}, w_{22}^{[1]}, w_{13}^{[1]}, w_{23}^{[1]}$ and those between hidden layer and output layer are $w_{11}^{[2]}, w_{21}^{[2]}, w_{31}^{[2]}$

Find the numerical value of $\frac{\partial \mathcal{L}}{\partial w_{12}^{[1]}}$. Answer upto 4 decimal places.

Consider a deep MLP and shallow MLP. Both gives the same loss and accuracy on the training data trained with the same number of samples.

(A) We prefer deep MLP (since deep neural networks are the best as of now)

(B) We prefer shallow MLP

(C) Both are equally good.

(D) Both neural networks then represent the same function. (since the loss is equal on both)

(E) None of the above.

Question Description

Quiz 3, Question 7

Question Image

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

The number of binary classifiers in a DDAG classifier with K classes to be evaluated at the test time *will be less/more/equal than K*

Question Description

Quiz 3, Question 3

Question Image

Remember the SVM problem from the problems we solved in the class. (1D samples)

$$(-1,+1),(0,-1),(+1,-1)$$

we geometrically solved the problem and saw the optimal primal solution as $w=-2$ and $b=-1$

Assume the samples were

$$(-1,-1),(0,+1),(+1,+1)$$

geometrically solve and give the answer as w=——,b=——

Question Image

Consider an MLP with 4 inputs, two hidden layers of 5 neurons each and one output neuron. All neurons have sigmoid activation. No bias.

How many learnable parameters are there in this network?

Question Description

Quiz 3, Question 5

Question Image

Consider an MLP with two input, one output and one hidden layer with two neurons. No bias. All weights are 1.0.

Hidden neurons have ReLu Activation and output has tanh activation.

Find the output of this MLP for an input of $[-2, -3]^T$

For Kernel Percepron

(A) It can be used for linearly separable or non-separable data

(B) At test time, we evaluate it as:

$$sign(\mathbf{w}^T\mathbf{x})$$

(C) At the test time, we evaluate it as:

$$sign(\sum_{i=1}^{N} \alpha_i y_i \kappa(\mathbf{x}_i, \mathbf{x})$$

(D) At the test time, we evaluate it as:

$$sign(\sum_{i=1}^{N} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x})$$

(E) when kernel is linear kernel, Kernel Perceptron reduces to the regular Perceptron.

Consider the popular activation function ReLu.

(A) its gradient can be either positive or. negative.

(B) its value can be either positive or negative

(C) it is an increasing function.

(D) it is a non-decreasing function

(E) all the above

Consider the popular activation function Leaky-ReLu.

(A) its gradient can be either positive or. negative.

(B) its value can be either positive or negative

(C) it is an increasing function.

(D) it is a non-decreasing function

(E) all the above

Question Details

Instructions Quiz

Type Multiple choice Questions (MCQ)

Question Quiz 3, Question 13

Click to zoom.

Consider an MLP with one hidden layer. $\mathbf{x}$ is the input and $\mathbf{y}$ is the output. All neurons in the hidden and output have ReLU activation.

(A) This network can be reduced to $\mathbf{y} = \mathbf{W}\mathbf{x}$

(B) This network can be modelled as: "Either $\mathbf{y} = \mathbf{W}_1\mathbf{x}$ or $\mathbf{y} = \mathbf{W}_2\mathbf{x}$"

(C) If all elements of $\mathbf{x}$ are negative, $\mathbf{y} = \mathbf{0}$.

(D) If $\mathbf{y} = \mathbf{0}$ imply that at least some of the elements of $\mathbf{x}$ are negative.

(E) None of the above.

Quiz

Multiple choice Questions (MCQ)

Quiz 3, Question 14

Consider an MLP which is getting trained with Back Propagation for a multiclass classification problem.

(A) The performance of the final model will depend on the initialization.

(B) The performance of the final model will depend on the learning rate we use.

(C) The performance of the final model will depend on the termination criteria we use.

(D) The performance of the final model will depend on the loss function we use.

(E) Exactly three of the above four are correct.

Consider a two class classification problem in 2-dimension with 6 data points.

$$\mathcal{D} = \{([0,0]^T, -), ([1,0]^T, -), ([0,1]^T, -), ([1,1], +), ([2,2]^T, +), ([2,0]^T, +)\}$$

We construct a hard margin SVM solution for this problem.

(A) Addition of $([0,2]^T, +)$ will change the support vector set, but not the margin.

(B) Addition of $([0, \frac{3}{2}]^T, +)$ will change the support vector set, and the margin.

(C) Addition of no sample can increase the margin.

(D) Addition of $([1,2]^T, +)$ does not change the support vector set and the margin.

(E) Addition of $([0, \frac{3}{2}]^T, +)$ will change the support vector set, but the number of support vectors will not change.

Question Image

Consider a two class classification problem in 2-dimension with 6 data points.

$$\mathcal{D} = \{([0,0]^T, -), ([1,0]^T, -), ([0,1]^T, -), ([1,1], +), ([2,2]^T, +), ([2,0]^T, +)\}$$

We construct a hard margin SVM solution for this problem. The decision boundary is:

(A) $2x_1 + 2x_2 = 3$

(B) $-2x_1 - 2x_2 = 3$

(C) $2x_1 + 2x_2 = -3$

(D) $-2x_1 - 2x_2 = -3$

(E) None of the above.

If there are 10 classes, a DDAG based multi class classifier will require —— binary classifiers to build the. DDAG.

Question Image

Consider two quadratic kernels: $\kappa_1(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T\mathbf{q} + 1)^2$ and $\kappa_2(\mathbf{p}, \mathbf{q}) = (\mathbf{p}^T\mathbf{q})^2$.

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

$\kappa_3() = \kappa_1() + \kappa_2()$ is also a valid kernel.

Consider a deep MLP and shallow MLP. Both are trained with the same number of samples.

(A) It is highly likely that Deep MLP will have lower training error. (since deeper the powerful!)

(B) It is highly likely that the shallow MLP will have lower training error. (since Occam's Razor says so)

(C) If the number of training samples is small, Deep MLP is going to overfit.

(D) If the number of training samples is small, Shallow MLP is going to overfit.

(E) None of the above

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

ReLu is a *piece-wise linear* **activation function.**

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

While training, an MLP with a hinge loss *solves a convex optimization problem.*

Make the necessary minimal changes (if any required) and rewrite as true sentences in the space provided. Avoid changing the words in bold.

The optimal solution to PCA and LDA *are always orthogonal.*

Consider a set of N valid kernels $\kappa_i(\cdot, \cdot)$

(A) $\sum_{i=1}^{N} \kappa_i()$ is also a valid kernel.

(B) $\sum_{i=1}^{N} \alpha_i \kappa_i()$ is also a valid kernel for any $\alpha_i \in R$.

(C) $\sum_{i=1}^{N} \alpha_i \kappa_i()$ is also a valid kernel for any $\alpha_i \in R^+$.

(D) $\Pi_{i=1}^{N} \kappa_i()$ is also a valid kernel.

(E) All the above.