

Intro into AI

Day 2 (Revision)

In supervised learning, what does it mean for a model to “generalize”?

- A. It achieves minimal loss on the training dataset
- B. It learns parameters that minimize the expected loss on unseen data
- C. It perfectly reconstructs the input–output mapping in the dataset
- D. It converges to a global minimum of the loss function

In supervised learning, what does it mean for a model to “generalize”?

- A. It achieves minimal loss on the training dataset
- B. It learns parameters that minimize the expected loss on unseen data**
- C. It perfectly reconstructs the input–output mapping in the dataset
- D. It converges to a global minimum of the loss function

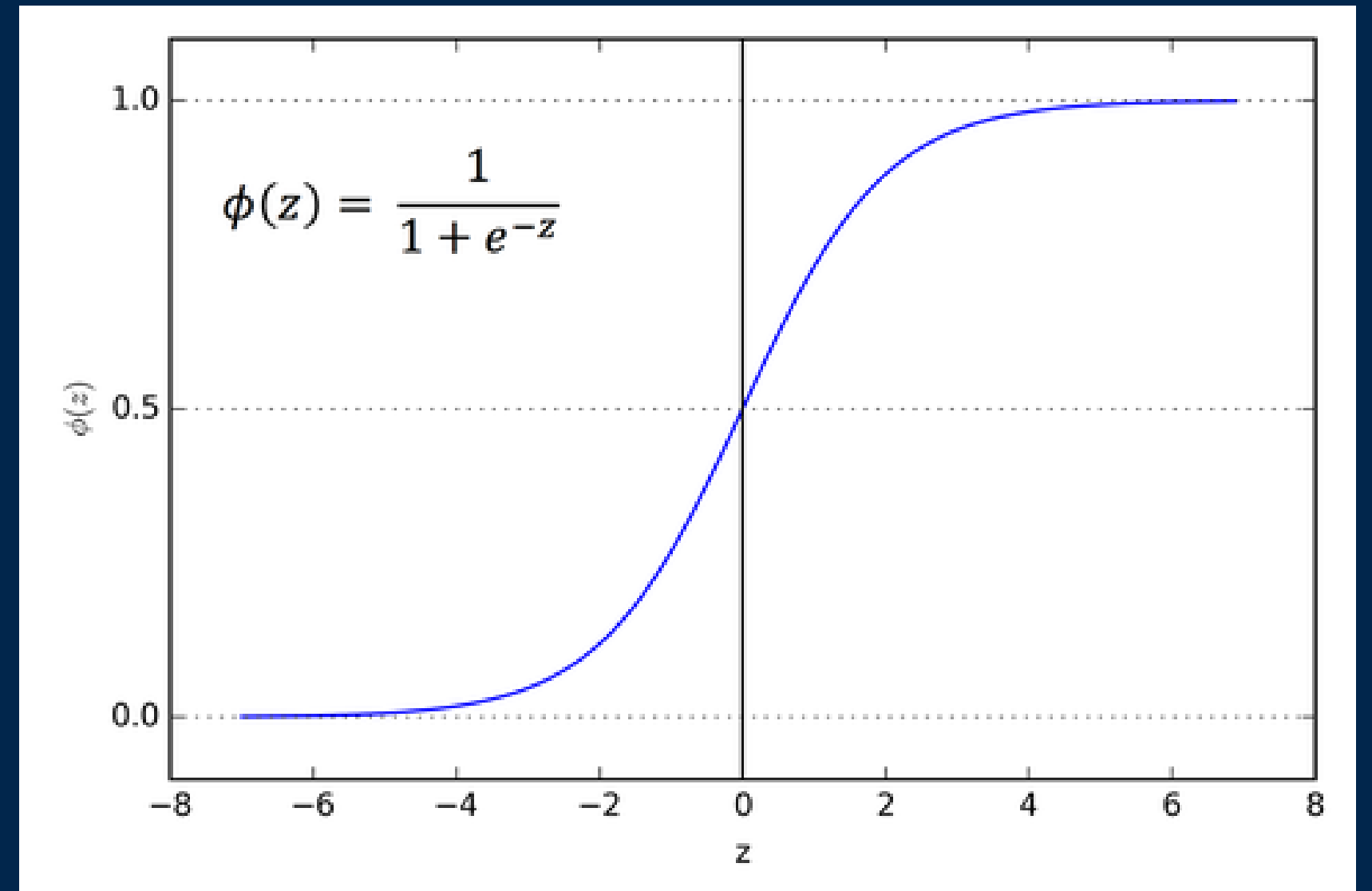
This formula below represents which function?

A. Linear Function

B. Sigmoid Function

C. Softmax Function

D. Polynomial Function



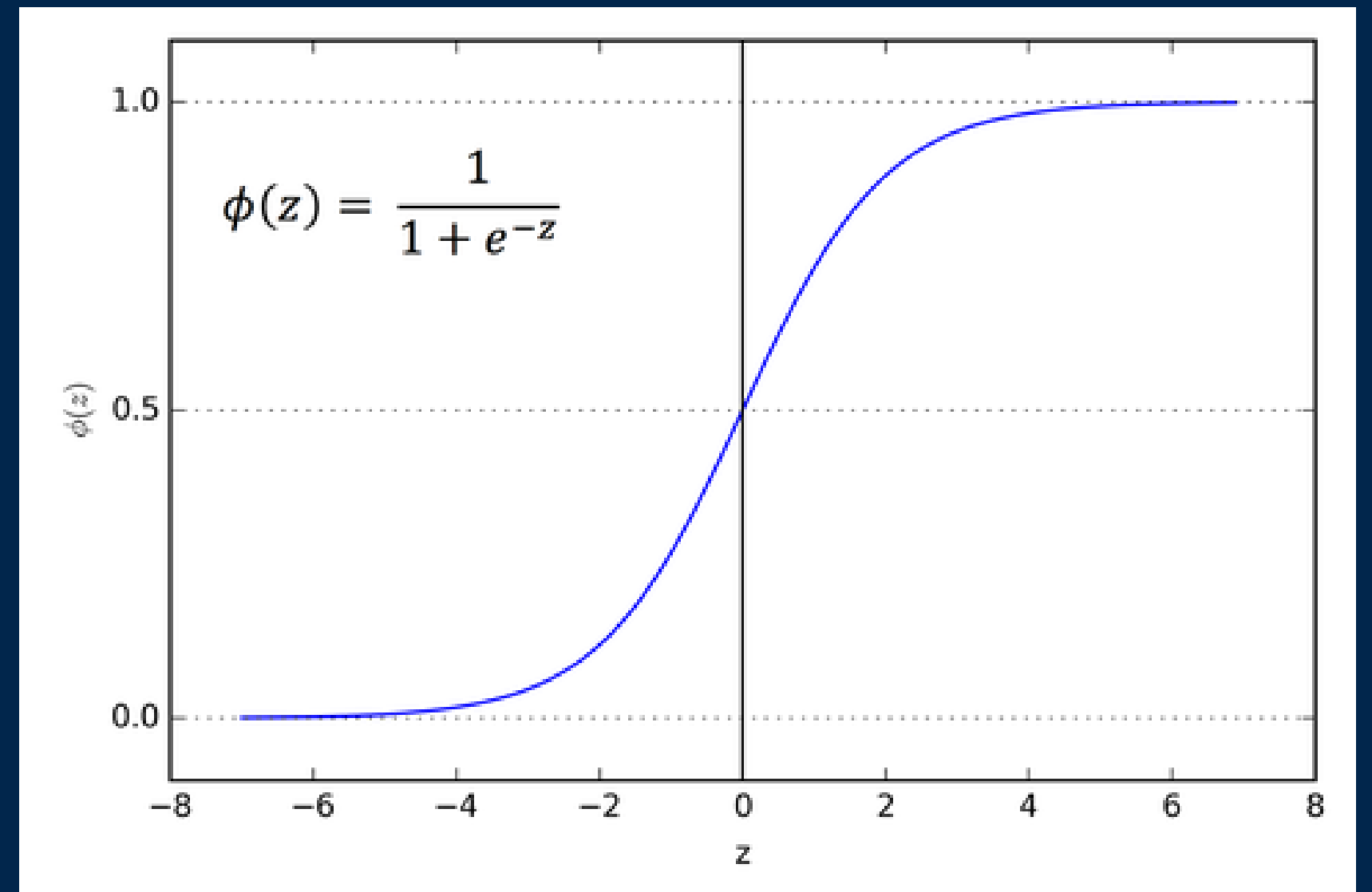
This formula below represents which function?

A. Linear Function

B. Sigmoid Function

C. Softmax Function

D. Polynomial Function



Which property of Mean Squared Error (MSE) most directly explains its sensitivity to outliers?

- A. It is convex with respect to model parameters
- B. It aggregates errors across all samples
- C. It squares the residual magnitude
- D. It is differentiable everywhere

Which property of Mean Squared Error (MSE) most directly explains its sensitivity to outliers?

- A. It is convex with respect to model parameters
- B. It aggregates errors across all samples
- C. It squares the residual magnitude**
- D. It is differentiable everywhere

In BCE loss, what causes a prediction with high confidence in the wrong class to incur a large penalty?

- A. The normalization by sample size
- B. The subtraction of true and predicted labels
- C. The logarithmic transformation of predicted probabilities
- D. The use of a threshold during classification

In BCE loss, what causes a prediction with high confidence in the wrong class to incur a large penalty?

- A. The normalization by sample size
- B. The subtraction of true and predicted labels
- C. The logarithmic transformation of predicted probabilities**
- D. The use of a threshold during classification

2 linear models w/ the same MSE. One has very large weights, the other small weights. Why is the second model preferred?

- A. Smaller weights reduce computational cost
- B. Smaller weights imply fewer features were used
- C. Smaller weights reduce sensitivity to small input changes
- D. Smaller weights guarantee convex optimization

2 linear models w/ the same MSE. One has very large weights, the other small weights. Why is the second model preferred?

- A. Smaller weights reduce computational cost
- B. Smaller weights imply fewer features were used
- C. Smaller weights reduce sensitivity to small input changes**
- D. Smaller weights guarantee convex optimization

Why does Lasso (L1) regularization often perform feature selection automatically?

- A. It penalizes squared weights
- B. Its constraint geometry has sharp corners
- C. It requires iterative optimization
- D. It minimizes variance more aggressively

Why does Lasso (L1) regularization often perform feature selection automatically?

- A. It penalizes squared weights
- B. Its constraint geometry has sharp corners**
- C. It requires iterative optimization
- D. It minimizes variance more aggressively

Why do decision trees not use gradient descent for learning splits?

- A. Their loss function is non-convex
- B. Their parameters are discrete
- C. Their splits are non-differentiable decisions
- D. Their objective function is undefined

Why do decision trees not use gradient descent for learning splits?

- A. Their loss function is non-convex
- B. Their parameters are discrete
- C. Their splits are non-differentiable decisions**
- D. Their objective function is undefined

What is the fundamental reason Random Forests reduce variance compared to single trees?

- A. Trees are trained sequentially
- B. Trees are trained on correlated subsets
- C. Predictions are averaged across independent models
- D. Predictions are averaged across independent models

What is the fundamental reason Random Forests reduce variance compared to single trees?

- A. Trees are trained sequentially
- B. Trees are trained on correlated subsets
- C. Predictions are averaged across independent models**
- D. Predictions are averaged across independent models

Why is feature scaling important for distance-based models like k-NN?

- A. To ensure convexity of the loss
- B. To reduce training time
- C. To allow probabilistic interpretation
- D. To stop features with large magnitudes from dominating distance calculation

Why is feature scaling important for distance-based models like k-NN?

- A. To ensure convexity of the loss
- B. To reduce training time
- C. To allow probabilistic interpretation
- D. To stop features with large magnitudes from dominating distance calculation**

Output: [0.1, 0.7, 0.2] for [Dog, Cat, Bird]. Which class is predicted?

A. Dog

B. Cat

C. Bird

D. None of the above

Output: [0.1, 0.7, 0.2] for [Dog, Cat, Bird]. Which class is predicted?

A. Dog

B. Cat

C. Bird

D. None of the above

k-NN has low training cost but high inference cost.

A. True

B. False

k-NN has low training cost but high inference cost.

A. True

B. False

Using a simple Linear Model for complex image pixels often leads to:

- A. Perfect accuracy
- B. Overfitting
- C. High complexity
- D. Underfitting

Using a simple Linear Model for complex image pixels often leads to:

A. Perfect accuracy

B. Overfitting

C. High complexity

D. Underfitting

Which is a defining characteristic of Deep Learning compared to traditional ML?

- A. Requires manual feature engineering
- B. Best for structured table data
- C. Automated feature extraction
- D. Lightweight and interpretable

Which is a defining characteristic of Deep Learning compared to traditional ML?

- A. Requires manual feature engineering
- B. Best for structured table data
- C. Automated feature extraction**
- D. Lightweight and interpretable

Why does increasing the regularization parameter λ too much lead to underfitting?

- A. It removes noise from the data
- B. It forces weights toward zero, reducing model flexibility
- C. It increases variance
- D. It makes the model memorize the training data

Why does increasing the regularization parameter λ too much lead to underfitting?

A. It removes noise from the data

B. It forces weights toward zero, reducing model flexibility

C. It increases variance

D. It makes the model memorize the training data

What does the sigmoid function primarily enable in logistic regression?

- A. Linear separation of data
- B. Mapping real-valued scores to probabilities
- C. Reducing variance
- D. Feature normalization

What does the sigmoid function primarily enable in logistic regression?

- A. Linear separation of data
- B. Mapping real-valued scores to probabilities**
- C. Reducing variance
- D. Feature normalization

Softmax guarantees that predicted class probabilities sum to 1.

A. True

B. False

Softmax guarantees that predicted class probabilities sum to 1.

A. True

B. False

Why are Support Vectors critical in SVMs?

- A. They determine the learning rate
- B. They are the farthest points from the boundary
- C. They define the maximum-margin hyperplane
- D. They eliminate the need for regularization

Why are Support Vectors critical in SVMs?

- A. They determine the learning rate
- B. They are the farthest points from the boundary
- C. They define the maximum-margin hyperplane**
- D. They eliminate the need for regularization

Gradient Boosting reduces error by training each new tree independently of the previous ones.

A. True

B. False

Gradient Boosting reduces error by training each new tree independently of the previous ones.

A. True

B. False

Which loss should be used for spam vs not spam classification?

- A. Mean Squared Error
- B. Categorical Cross-Entropy
- C. Hinge Loss
- D. Binary Cross Entropy (Log Loss)

Which loss should be used for spam vs not spam classification?

- A. Mean Squared Error
- B. Categorical Cross-Entropy
- C. Hinge Loss
- D. Binary Cross Entropy (Log Loss)**

Which loss is correct for digit classification (0–9), one digit per image?

A. Binary Cross Entropy (Log Loss)

B. Categorical Cross-Entropy

C. Mean Absolute Error

D. Hinge Loss

Which loss is correct for digit classification (0–9), one digit per image?

A. Binary Cross Entropy (Log Loss)

B. Categorical Cross-Entropy

C. Mean Absolute Error

D. Hinge Loss

What does a model output in multiclass classification?

- A. One probability per sample
- B. A probability distribution over K classes
- C. A binary label
- D. Independent class scores

What does a model output in multiclass classification?

- A. One probability per sample
- B. A probability distribution over K classes**
- C. A binary label
- D. Independent class scores

Which of the following is considered a deep learning model?

A. MLP

B. K-NN

C. Linear Regression

D. Decision Tree

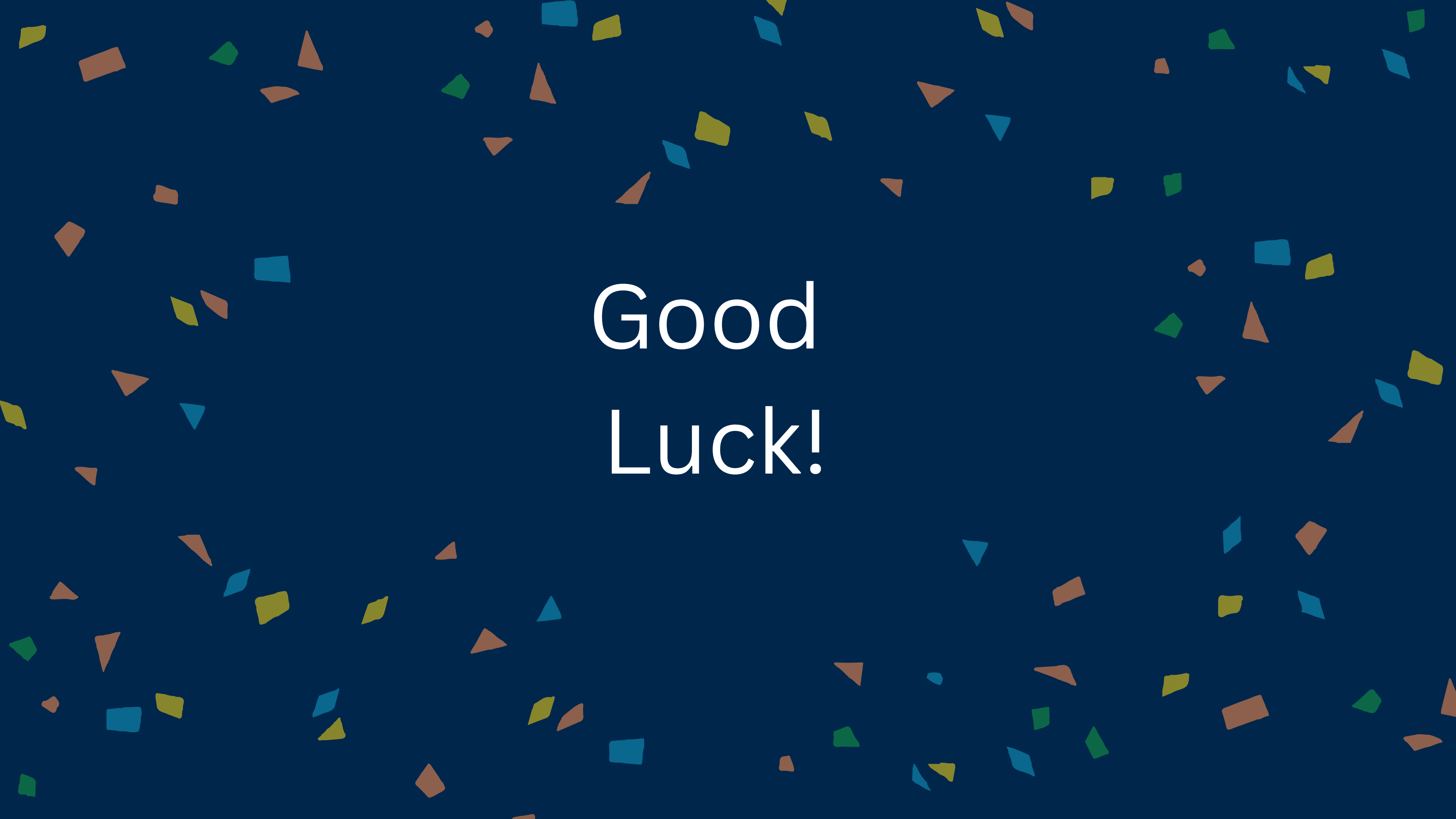
Which of the following is considered a deep learning model?

A. MLP

B. K-NN

C. Linear Regression

D. Decision Tree



Good
Luck!