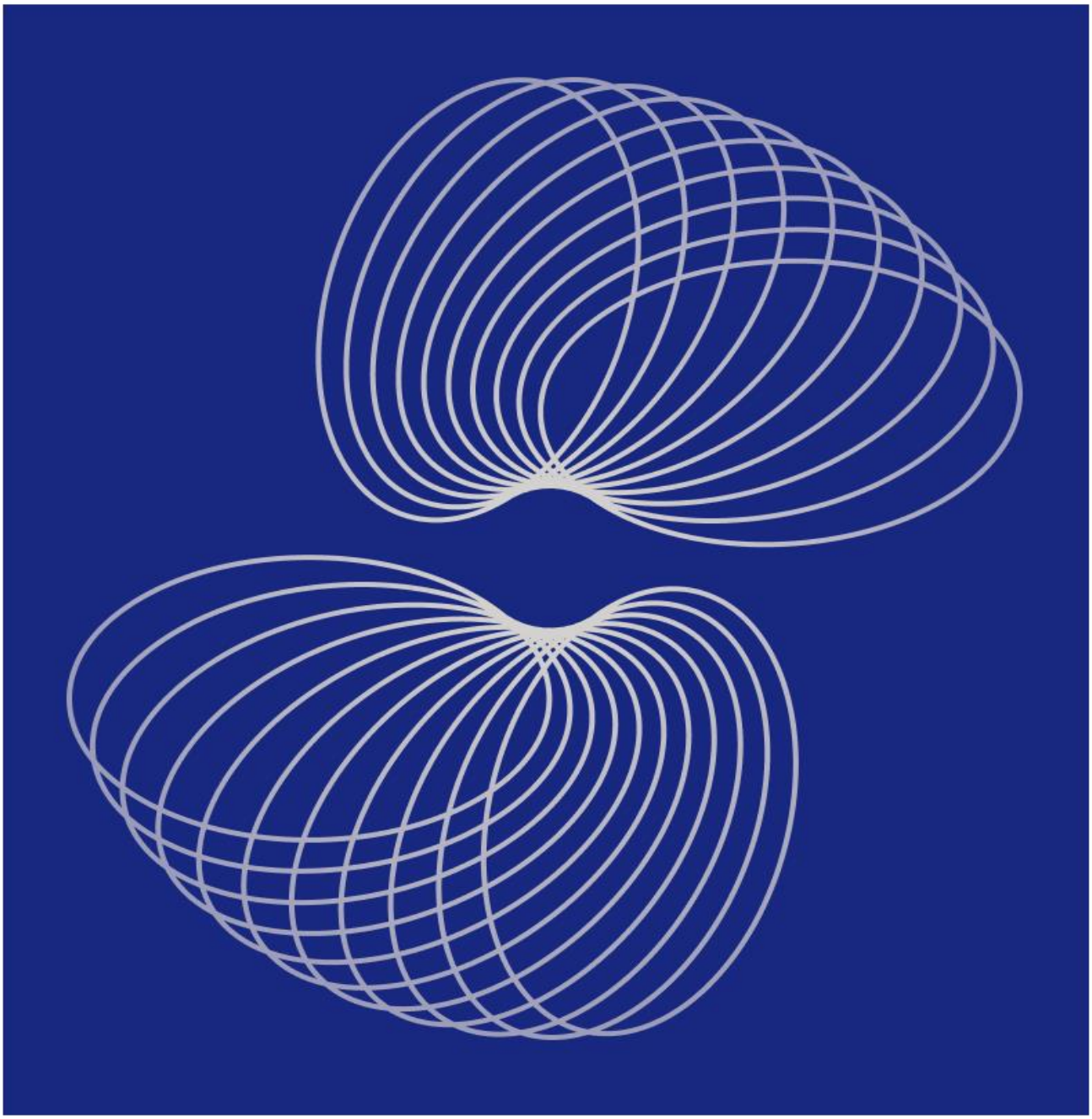
Univ.AI

AI Quizbook

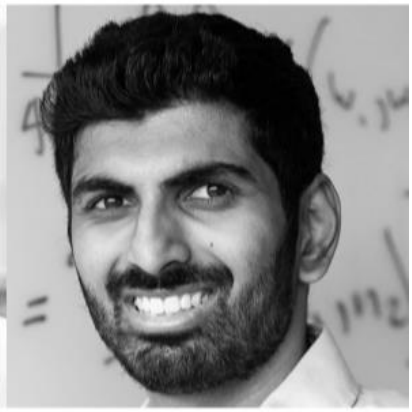
2020 | Issue 2



Founded by Harvard Professors to bring you the world's best AI/ML training

Rahul Dave
Chief Scientist,
Univ.AI
Former Lecturer,
Harvard University

Pavlos Protopapas
Scientific Program
Director, IACS,
Harvard University



Achuta Kadambi
Asst. Prof. UCLA
PhD, MIT Media Lab

All Programs Live! Online

Elite Faculty from Top Universities

Cutting-Edge Curriculum

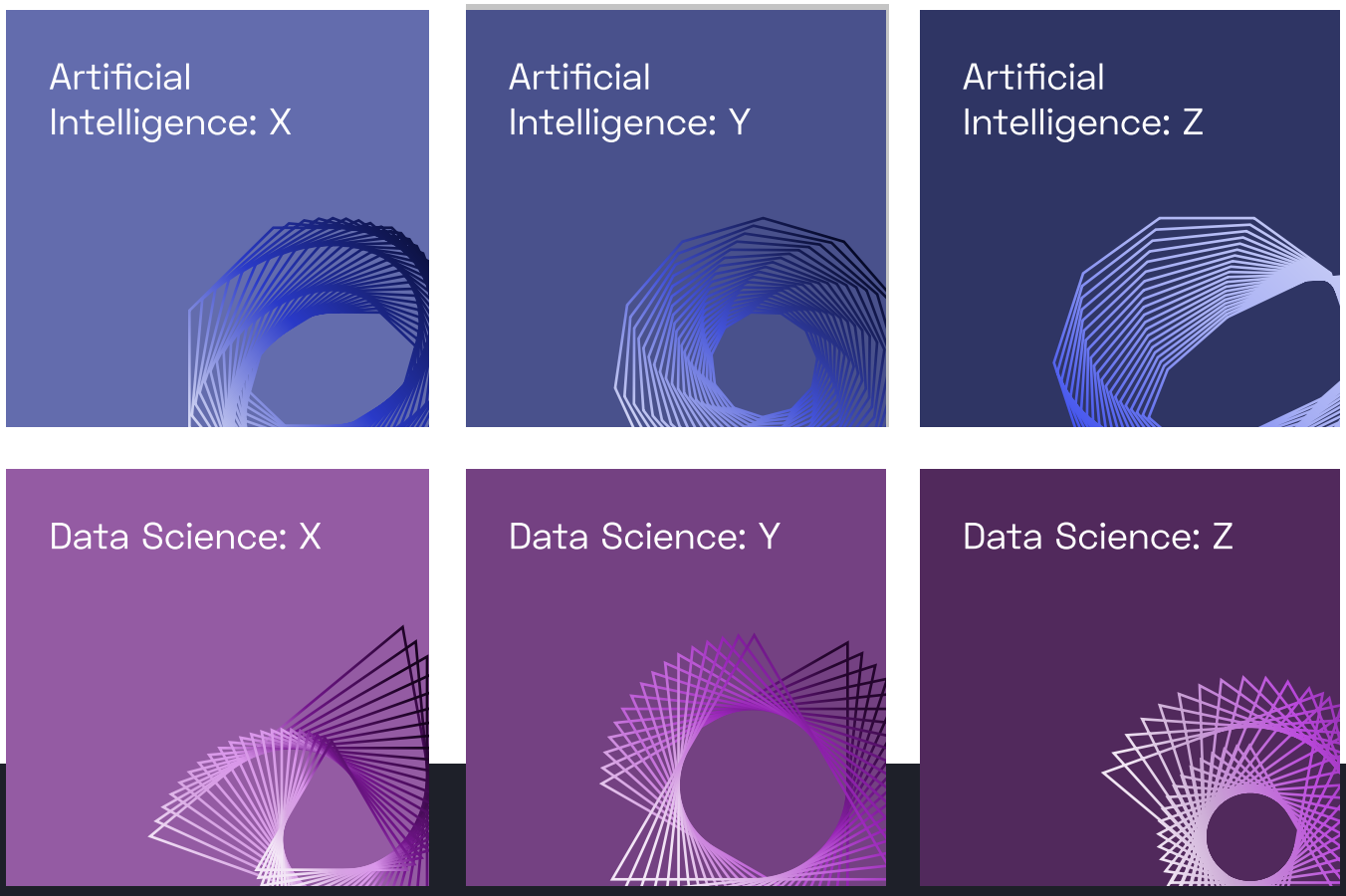
Live! Hands-On Mentorship

Learning by Lots of Doing!

Train to be among the best

Univ.AI offers you a 3-step path to a GUARANTEED high-end job in Machine Learning & Artificial Intelligence, Data Science, or Programming (Coming Soon)

- Learn LIVE-online with Harvard faculty
- Gain expertise in 3 sequential, standalone certifications of 10-12 weeks each
- Get job-ready outcomes and certifications at each level
- The beginner programs are offered at very low costs
- Advance to the Master Certification, risk your investment with a state-of-the-art employment guarantee



Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of n classes (p1, p2..pk) such that sum of p over all n equals to 1?

- [ ] Softmax
- [ ] ReLU
- [ ] Sigmoid
- [ ] Tanh

Answer on the next page

Which of the following statements is the best description of early stopping?

- Train the network until a local minimum in the error function is reached
- Form of regularization used to avoid overfitting. Stops training when the generalization error starts to increase
- Add a momentum term to the weight update in the Generalized Delta Rule, so that training converges more quickly
- A faster version of backpropagation, such as the `Quickprop' algorithm

Previous Answer
Softmax

The number of nodes in the input layer is 15 and the hidden layer is 4. The maximum number of connections from the input layer to the hidden layer are

- [ ] 60
- [ ] Less than 60
- [ ] More than 60
- [ ] It is an arbitrary value

Previous Answer
Form of regularization used to avoid overfitting. Stops training when the generalization error starts to increase

Which of the following is/are (an) important step(s) to pre-process the text in NLP (Natural Language Processing) based projects?

- [ ] Stemming
- [ ] Stop word removal
- [ ] Object Standardization
- [ ] All of the above

Previous Answer
60

What would you do in PCA (Principal Component Analysis) to get the same projection as SVD (Singular Value Decomposition)?

- [ ] Transform data to zero mean
- [ ] Transform data to zero median
- [ ] Not possible
- [ ] None of the above

Previous Answer
All of the above

Which of the following is an example of feature extraction?

- [ ] Constructing a word vector from an email
- [ ] Applying PCA projects to a large high-dimensional data
- [ ] Using Independent Component Analysis to differentiate speakers in an audio recording
- [ ] All of the above

Previous Answer
Transform data to zero mean

Which of the following statements is true when you use 1×1 convolutions in a CNN?

- [ ] It can help in dimensionality reduction
- [ ] It can be used for feature pooling
- [ ] It suffers less overfitting due to small kernel size
- [ ] All of the above

Previous Answer
All of the above

Which of the following statements is true about output of a dropout layer with 0.8 as keep probability and n as the input size of the dropout layer (and thus the output size of the previous layer)?

- [ ] Output size will be exactly 0.8*n
- [ ] Output size will be approximately 0.8*n
- [ ] Output size of dropout layer will be n with roughly 0.2*n output activations from the previous layer randomly set to 0 in the training phase
- [ ] Output size will be same as n with exactly 0.2*n entries converted to 0

Previous Answer
All of the above

Each point in hierarchical clustering is part of

- [ ] Single cluster
- [ ] Many clusters
- [ ] None of the above

Previous Answer
Output size of dropout layer will be n with roughly 0.2*n output activations from the previous layer randomly set to 0 in the training phase

What are the steps for using a gradient descent algorithm?

a. Calculate error between the actual and predicted value
b. Reiterate until you find the best weights of the network
c. Pass an input through the network and get values from the output layer
d. Initialize random weights and biases
e. Change the weights of each neuron to reduce the error

- [ ] d, c, e, a, b
- [ ] e, d, b, c, a
- [ ] a, c, b, e, d
- [ ] d, c, a, e, b

Previous Answer
Many clusters

Given spam emails are the positive class, how many of the actual spam emails are correctly predicted as spam?

- [ ] (True Positives)/(True Positives + False Negatives)
- [ ] (True Positives)/(True Positives + True Negatives)
- [ ] (True Positives)/(False Positives + True Negatives)
- [ ] (True Positives)/(False Positives + False Negatives)

Previous Answer
d, c, a, e, b

Which of the following is a disadvantage of decision trees?

- [ ] Training time is relatively high as complexity is low
- [ ] Decision trees are robust to outliers
- [ ] Decision trees are prone to be overfit
- [ ] None of the above

Previous Answer
(True Positives)/(True Positives + False Negatives)

Which of the following components make a neural network nonlinear in nature?

- [ ] Hidden Layers
- [ ] Activation functions
- [ ] Weights and Biases
- [ ] Regularization

Previous Answer
Decision trees are prone to be overfit

How many input nodes are required to process a grayscale image of 28X28?

- [ ] 28
- [ ] 784
- [ ] 2354
- [ ] 56

Exploding/vanishing gradients are problematic to a greater extent (due to a large network size obtained by unrolling the network through time) in the case of X Neural Networks. What is X?

Previous Answer
784

What is/are true about ridge regression?

1. When lambda is 0, model resembles a linear regression model
2. When lambda is 0, model doesn't resemble a linear regression model
3. When lambda goes to infinity, we get very small coefficients approaching 0
4. When lambda goes to infinity, we get very large coefficients approaching infinity

- [ ] 1 and 3
- [ ] 1 and 4
- [ ] 2 and 3
- [ ] 2 and 4

X is used for summarizing the performance of a classification algorithm and contains cells indicating some errors such as “true positives” and “false negatives.” What is X?

Previous Answer
1 and 3

Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

- [ ] Decision Tree
- [ ] Regression
- [ ] Classification
- [ ] Random Forest

The statement “L1 corresponds to setting a Laplacian prior on the terms, while L2 corresponds to a Gaussian prior” about L1 and L2 regularization is true or false?

- [ ] True
- [ ] False

Previous Answer
Random Forest

What cross-validation technique would be best suited for a time series dataset?

- [ ] Standard k-folds cross-validation
- [ ] Cross-validation in which the test set is always later in time than the training and validation sets
- [ ] Time series datasets do not require cross validation as the data is not randomly distributed
- [ ] None of the above

Previous Answer
True

Univ.AI

To learn more, drop us a note at hello@univ.ai

Previous Answer

Cross-validation in which the test set is always later in time than the training and validation sets