edquest

MLOps Training

MACHINE LEARNING OPERATIONS

edquest

Day 3 Agenda

Regularization & Deep Learning

- Recap
- Quiz for the previous Day
- Overfitting, Underfitting
- Workshop
- L1/L2 regularization
- Intro to neural networks and deep learning flow
- Quiz

edquest

Recap

EXPLORATORY DATA ANALYSIS (EDA)

edquest

Q1: The syntax df.shape returns a tuple representing ___ of the dataset.

A. correlation and variance

B. mean and median

C. number of rows and columns

D. size in memory

edquest

Q1: The syntax df.shape returns a tuple representing ___ of the dataset.

A. correlation and variance

B. mean and median

C. number of rows and columns

D. size in memory

edquest

Q2: To check the percentage of missing values column-wise, we use df.isnull().mean(), which gives us values between ____ and ____.

A. 1 and 100

B. 0 and 1

C. -1 and 1

D. 0 and 100

edquest

Q2: To check the percentage of missing values column-wise, we use df.isnull().mean(), which gives us values between ___ and ___.

A. 1 and 100

B. 0 and 1

C. -1 and 1

D. 0 and 100

edquest

Q3: The line

df = df.drop(columns=null_percent[null_percent > 0.3].index) is used to remove columns with more than ____% missing values.

A. 3

B. 10

C. 30

D. 50

edquest

Q3: The line

df = df.drop(columns=null_percent[null_percent > 0.3].index) is used to remove columns with more than ____% missing values.

A. 3

B. 10

C. 30

D. 50

edquest

Q4: To examine the summary statistics of numerical features, the code `df[num_cols].describe().T[:5]` is used. The .T stands for ____.

A. table format

B. trimming values

C. transpose

D. top values

edquest

Q4: To examine the summary statistics of numerical features, the code df[num_cols].describe().T[:5] is used. The .T stands for ____.

A. table format

B. trimming values

C. transpose

D. top values

edquest

Q5: The function sns.boxplot() is used to detect ___ and visualize distribution characteristics.

A. duplicates

B. outliers

C. nulls

D. categories

edquest

Q5: The function sns.boxplot() is used to detect ____ and visualize distribution characteristics.

A. duplicates

B. outliers

C. nulls

D. categories

edquest

Q6: A histogram with KDE (Kernel Density Estimation) is plotted using `sns.histplot(df['SalePrice'], kde=True)`. The KDE line helps to understand the ___ of the data.

A. completeness

B. sparsity

C. central tendency

D. probability density

edquest

Q6: A histogram with KDE (Kernel Density Estimation) is plotted using `sns.histplot(df['SalePrice'], kde=True)`. The KDE line helps to understand the ___ of the data.

A. completeness

B. sparsity

C. central tendency

D. probability density

edquest

Q7: The df.select_dtypes(include='object') line is used to select ___ features in the dataset.

A. numerical

B. categorical

C. Boolean

D. missing

edquest

Q7: The df.select_dtypes(include='object') line is used to select ___ features in the dataset.

A. numerical

B. categorical

C. Boolean

D. missing

edquest

Q8: The correlation matrix is created using df.corr(numeric_only=True). Setting numeric_only=True ensures that only ____ data types are used in the computation.

A. float and bool

B. object and float

C. int and float

D. datetime and object

edquest

Q8: The correlation matrix is created using df.corr(numeric_only=True). Setting numeric_only=True ensures that only ____ data types are used in the computation.

A. float and bool

B. object and float

C. int and float

D. datetime and object

edquest

Q9: The sns.heatmap() function with annot=True overlays ____ on the heatmap cells.

A. feature names

B. colors

C. correlation values

D. histogram bars

edquest

Q9: The sns.heatmap() function with annot=True overlays ____ on the heatmap cells.

A. feature names

B. colors

C. correlation values

D. histogram bars

edquest

Q10: In the context of EDA, the main reason to remove columns with high null percentages is to avoid ___ during model training.

A. faster execution

B. better styling

C. bias and error

D. overfitting

edquest

Q10: In the context of EDA, the main reason to remove columns with high null percentages is to avoid ___ during model training.

A. faster execution

B. better styling

C. bias and error

D. overfitting

edquest

Model’s Quality

Model’s Quality

edquest

Model’s Quality

A machine learning model is considered "good" when it performs well on both the training data and unseen (test) data.

What does that mean?

- Effectively captures meaningful patterns from the training dataset
- Performs reliably on new, unseen data through strong generalization
- Strikes a balance by avoiding both overfitting and underfitting

Regression Model Metrics

edquest

Regression

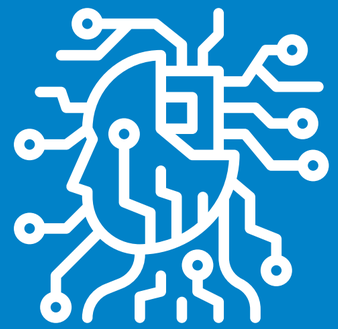
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2 Score (Coefficient of Determination)

Classification Model Metrics

edquest

Classification

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix



edquest

Bias & Variance

Bias

Bias

Bias refers to the error introduced when a machine learning model is overly simplistic and fails to capture the underlying patterns in the data.

Bias Example

If a model learns that every two-wheeler is a motor bike, it's making a high-bias assumption. In reality, a two-wheeler could be a scooty, moped, or even a bicycle.

This oversimplification leads to incorrect predictions — and that’s what bias is: the model makes generalized assumptions that miss important distinctions in the data.

Motor Bike

Motor Bike

Motor Bike

Bias Usecase

Medical Diagnosis

A model only uses basic symptoms (like fever or cough) to predict diseases.

It ignores more specific indicators (blood tests, scan reports).

Result: Predicts many patients incorrectly → high bias.

Bias Pros/Cons

High bias typically leads to underfitting, where the model performs poorly on both training and testing data because it fails to learn enough from the data.

Pro: It’s simple, fast, and less prone to overfitting.

Cons: It misses important patterns, leading to inaccurate predictions.

Variance

Variance

Variance is the error that happens when a machine learning model learns too much from the training data, including the noise, and fails to perform well on new data.

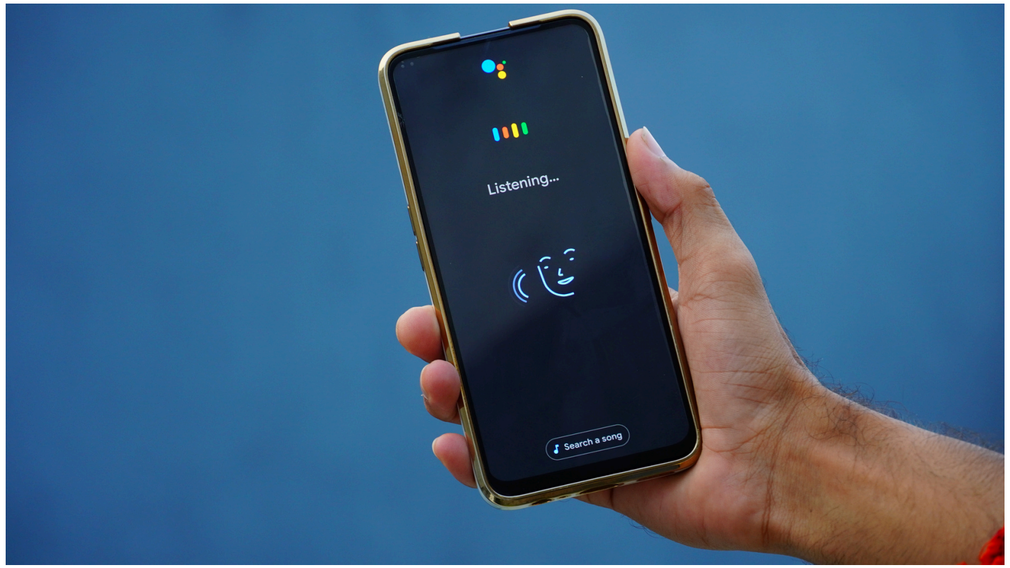
Variance Usecase

Voice Assistant Personalization

A voice assistant is trained heavily on one user's voice patterns.

When used by another person, it struggles to understand commands.

Result: Overfit to one voice → high variance.

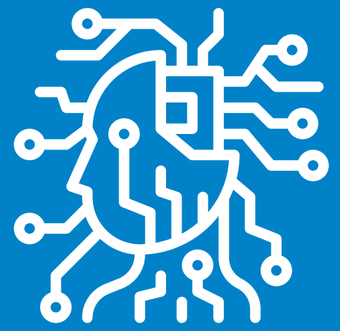


Variance Pros/Cons

High variance typically leads to overfitting, where the model performs well on training data but poorly on testing data.

Pro: It captures complex patterns in the data.

Con: It struggles to generalize to new, unseen data.



edquest

Underfitting & Overfitting

Underfitting

What is Underfitting ?

Underfitting occurs when the model hasn’t learned enough from the training data — either due to insufficient training time or because the input features lack the significance or volume needed to capture a meaningful relationship with the target variable.

Underfitting usecases

1. Housing price prediction
2. Weather forecasting
3. Image recognition
4. Movie Recommendation System
5. Speech Recognition System
6. Loan Approval Prediction

edquest

Reasons of Underfitting

- The model is too simple, so it can’t understand the real patterns in the data.
- The input features don’t give enough useful information to make good predictions.
- There’s not enough training data, so the model can’t learn well.
- Too much regularization is applied, which limits the model from learning properly.
- The features are not scaled, making it hard for the model to find correct relationships.

Overfitting

What is Overfitting?

The situation where any given model is performing too well on the training data but the performance drops significantly over the test set is called an overfitting model..

edquest

Overfitting usecases

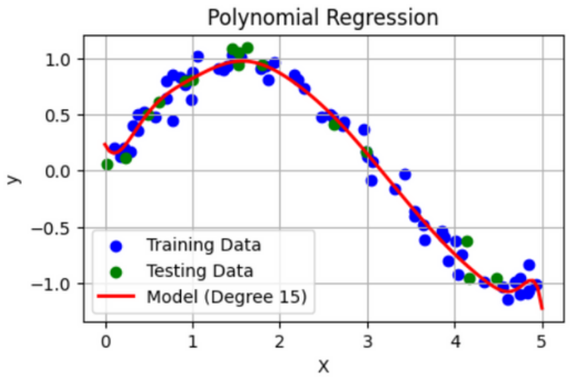
1. Online Ad Click Prediction
2. Credit Card Fraud Detection
3. Medical diagnosis model
4. Stock price prediction

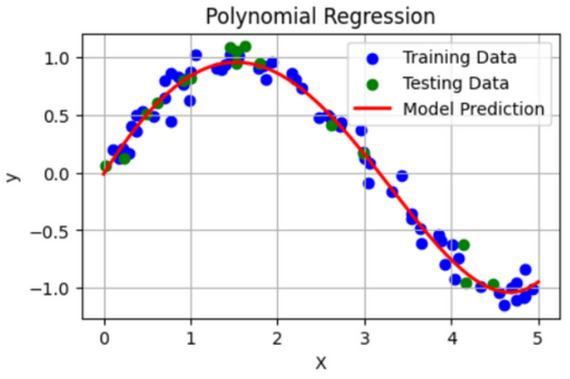
edquest

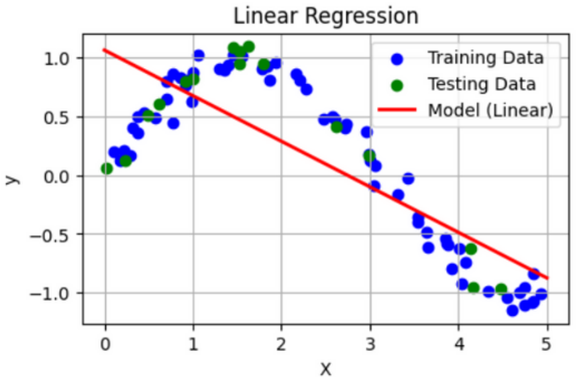
Reasons of Overfitting

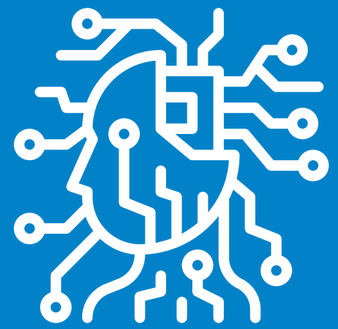
1. The model is too complicated and tries to learn everything, even the noise.
2. There is not enough training data, so the model just memorizes what it sees.
3. The model is very sensitive to the data, which means small changes confuse it easily.

Underfit & Overfit









edquest

Regularization

Need of Regularization

- Regression models can overfit when features are highly correlated or too many features exist.
- Overfitting leads to poor generalization.
- Regularization combats overfitting by penalizing large coefficients.

What is Regularization ?

- Regularization adds a penalty to the loss function
- Helps control model complexity
- Common in linear models: modifies OLS (Ordinary Least Squares)

Type of Regularization ?

- Ridge Regression or L2 Regularization
- Lasso Regression or L1 Regularization
- Elastic Net: Combining L1 & L2

edquest

Ridge Regression
or
L2 Regularization

- Adds squared magnitude of coefficients to loss.
- Penalizes large coefficients.
- Doesn't eliminate features but shrinks them.
- Best for multicollinearity.

edquest

Ridge Regression
or
L2 Regularization

```python
from sklearn.linear_model import Ridge

# alpha is the regularization strength
ridge = Ridge(alpha=1.0)
ridge.fit(X_train, y_train)
```

```python
from sklearn.linear_model import SGDRegressor

# Initialize SGDRegressor with L2 penalty
sgd_reg = SGDRegressor(penalty='l2')
sgd_reg.fit(X_train, y_train)
```

edquest

When to Use Ridge Regression ?

- You have **many correlated features**
- You want to **retain all**, but in controlled amounts

edquest

Lasso Regression
or
L1 Regularization

- Adds absolute value of coefficients to loss
- Can zero out coefficients — helps in feature selection
- Good when we believe only some features are relevant

When to Use Lasso Regression ?

- You need feature selection
- You suspect some features are noise

edquest

Elastic Net

- Mixes Ridge and Lasso penalties
- Helps when features are correlated and selection is needed
- Controlled by parameters **λ** (lambda) and **α** (alpha)

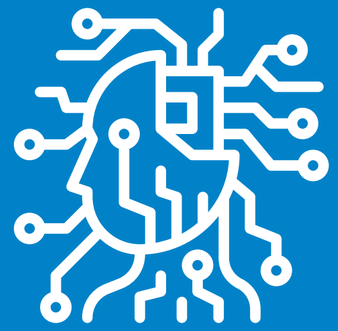
When to Use Elastic Net ?

- You want a **flexible balance**
- You’re unsure which approach is best

Regularization Summary

- Regularization reduces overfitting
- Ridge shrinks, Lasso selects, ElasticNet balances
- Tuning λ (lambda) and α (alpha) is essential
- Use cross-validation to pick best model

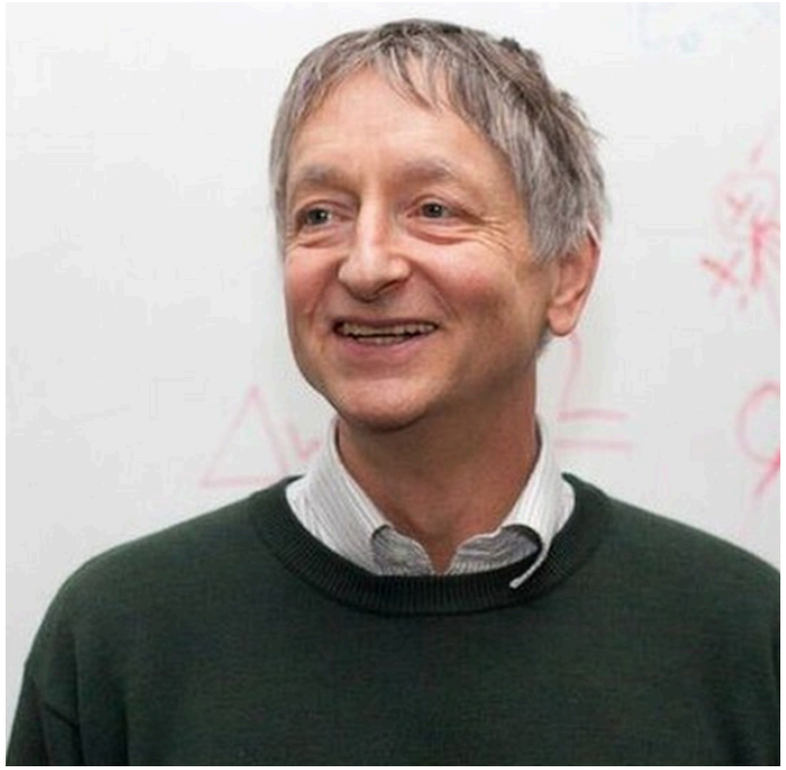
edquest



edquest

Deep Learning & Neural Network

Heroes of Deep Learning



Geoffery Hinton

Andrew NG

AI is the new electricity – By **Andrew NG**

Deep Learning

edquest

What is Deep Learning ?

- Subset of Machine Learning that mimics the **human brain**
- Uses neural networks with **multiple layers** ("deep" means many layers)
- Learns features **automatically** from raw data

Fact: The term “deep” in Deep Learning refers to the **number of layers** in a neural network.

Why Deep Learning now ?

- Explosion of data (Big Data)
- Increased compute power (GPUs, TPUs)
- Better algorithms (Activation functions, Optimizers)
- Open-source tools (TensorFlow, PyTorch)

Fact: Training a deep network that once took weeks can now be done in **minutes** using modern hardware.

edquest

Why Traditional ML is not enough ?

- Requires manual **feature engineering**
- Doesn’t scale well with **unstructured data** (images, audio, video)
- Performance plateaus as data grows

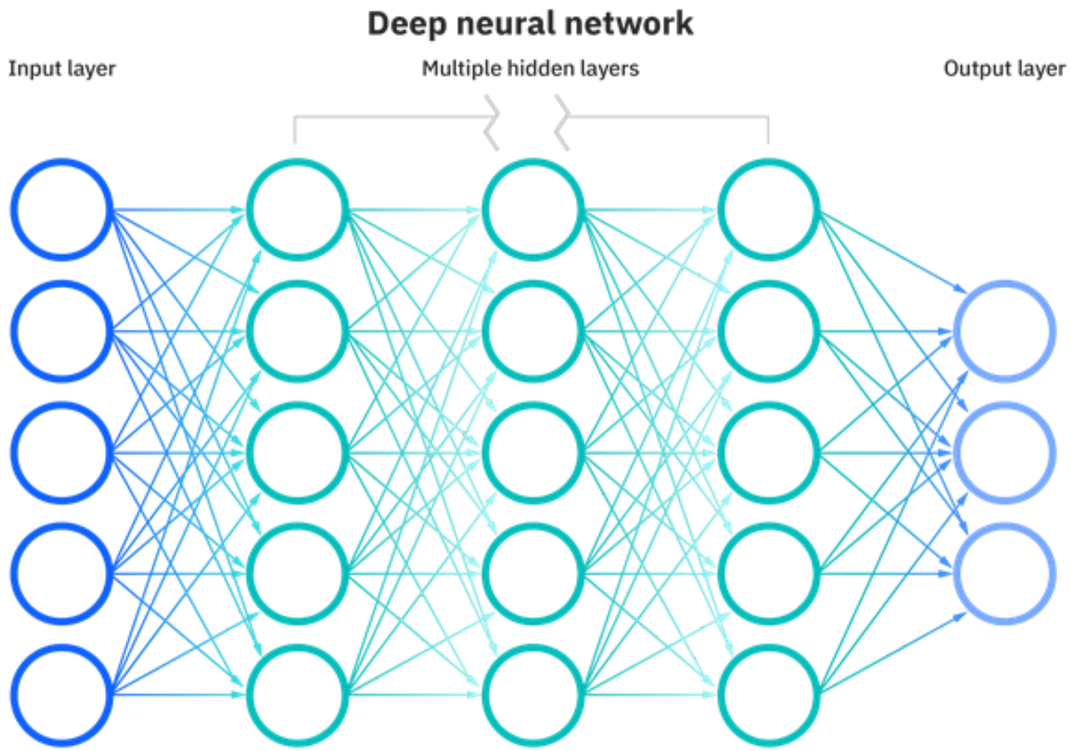
Examples:

- Predicting stock prices with regression
- Image classification using SVM

Fact: Deep learning **improves with more data**, while classical ML often hits a performance ceiling.

What is Neural Network ?

- Inspired by the human brain’s neurons
- Made up of **layers of nodes** (neurons): Input → Hidden → Output
- Each neuron applies a weighted sum and activation function



edquest

Deep Learning Applications

- Image Recognition (Face ID, Self-driving cars)
- Natural Language Processing (ChatGPT, Google Translate)
- Speech Recognition (Alexa, Siri)
- Healthcare (Disease detection from X-rays/MRIs)

edquest

Deep Learning Summary

- Deep learning is ideal for complex, high-dimensional data
- Neural networks can learn **hierarchical representations**
- Classical ML still has its place, but deep learning is driving the **future of AI**

edquest

edquest

Quiz

edquest

Q1. Overfitting occurs when a model performs well on training data but poorly on test data because it ____ the training patterns.

A. ignores

B. generalizes

C. memorizes

D. transforms

edquest

Q1. Overfitting occurs when a model performs well on training data but poorly on test data because it ____ the training patterns.

A. ignores

B. generalizes

C. memorizes

D. transforms

edquest

Q2. Underfitting typically happens when the model is too ____ to capture the underlying structure of the data.

A. large

B. simple

C. complex

D. deep

edquest

Q2. Underfitting typically happens when the model is too ____ to capture the underlying structure of the data.

A. large

B. simple

C. complex

D. deep

edquest

Q3. One of the key goals of L1 and L2 regularization is to prevent ____ by penalizing large weights.

A. optimization

B. dropout

C. overfitting

D. training

edquest

Q3. One of the key goals of L1 and L2 regularization is to prevent ____ by penalizing large weights.

A. optimization

B. dropout

C. overfitting

D. training

edquest

Q4. L1 regularization encourages ____ by forcing some weights to become exactly zero.

A. sparsity

B. overfitting

C. weight duplication

D. normalization

edquest

Q4. L1 regularization encourages ___ by forcing some weights to become exactly zero.

A. sparsity

B. overfitting

C. weight duplication

D. normalization

edquest

Q5. In L2 regularization, the penalty added to the loss function is proportional to the ___ of the weights.

A. mean

B. square

C. absolute value

D. maximum

edquest

Q5. In L2 regularization, the penalty added to the loss function is proportional to the ___ of the weights.

A. mean

B. square

C. absolute value

D. maximum

edquest

Q6. A neural network consists of input, hidden, and output layers where each neuron computes a weighted sum followed by a(n) ___ function.

A. loss

B. dropout

C. activation

D. optimizer

edquest

Q6. A neural network consists of input, hidden, and output layers where each neuron computes a weighted sum followed by a(n) ___ function.

A. loss

B. dropout

C. activation

D. optimizer

edquest

Q7. In deep learning, the term “deep” refers to the presence of ___ layers in the network.

A. very wide

B. many hidden

C. few output

D. multiple input

edquest

Q7. In deep learning, the term “deep” refers to the presence of ___ layers in the network.

A. very wide

B. many hidden

C. few output

D. multiple input

edquest

Q8. Regularization is typically added to the model’s ____ function to constrain the learning process.

A. activation

B. loss

C. dropout

D. optimizer

edquest

Q8. Regularization is typically added to the model’s ____ function to constrain the learning process.

A. activation

B. loss

C. dropout

D. optimizer

edquest

Q9. A key indicator of overfitting is when training accuracy is high, but ___ accuracy is low.

A. validation

B. learning rate

C. training

D. test loss

edquest

Q9. A key indicator of overfitting is when training accuracy is high, but ___ accuracy is low.

A. validation

B. learning rate

C. training

D. test loss

edquest

Q10. One benefit of using dropout layers in neural networks is that they prevent co-adaptation of neurons by randomly ___ some during training.

A. scaling

B. deleting

C. updating

D. deactivating

edquest

Q10. One benefit of using dropout layers in neural networks is that they prevent co-adaptation of neurons by randomly ___ some during training.

A. scaling

B. deleting

C. updating

D. deactivating