

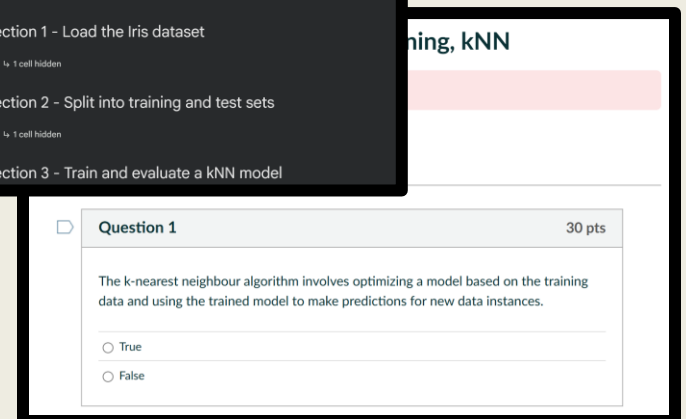
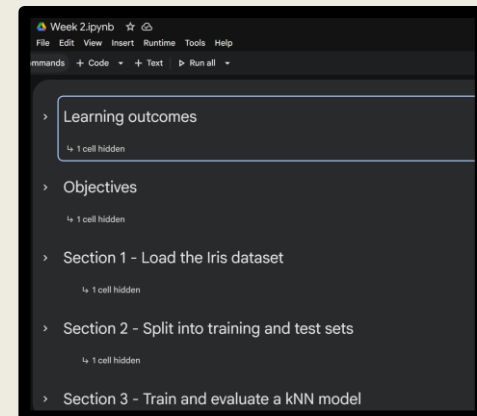
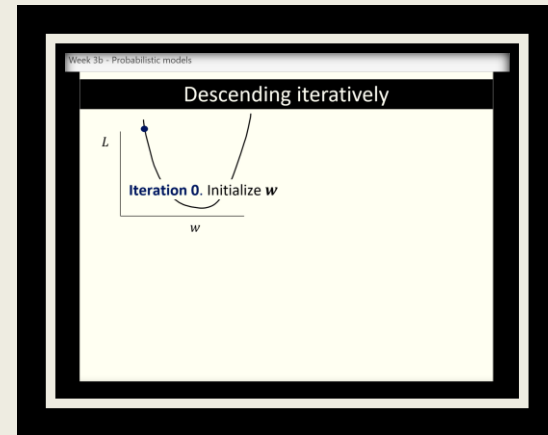
Week 2 discussion

MACHINE LEARNING

Dr. Temitayo Olugbade

Student To Do

- ❖ Watch Week 3 mini-videos ahead of next lecture
- ❖ Get to Week 2 code notebook & ungraded quiz when you can soon



Week 2 mini-video content

This week, you've been looking at:

- A tree-based model
- k-nearest neighbours

Learning outcomes



During this lecture, we'll explore together the questions below:

- ☐ Would differing feature scales affect the behaviour of the model?
- ☐ A student-posed question
- ☐ What does this AI do?

Discussion outline

- ❑ Would differing feature scales affect the behaviour of the model? **(40mins)**
- ❑ A student-posed question **(30mins)**
- ❑ What does this AI do? **(30mins)**



Discuss, Vote, Note

Temitayo Olugbade • 2h

Week 2 Student Post-Discussion Notes

Write a quick summary of your group's conclusions

Discussion questions

Pinned

Temitayo Olugbade... /teacher/
2 months ago

- How would differences in scales of different model input dimensions affect what a given ML model learns during training?
- What does this AI do?

Vote here

Pinned

Temitayo Olugbade... /teacher/
2 hours ago

Would differing scales across features affect the behaviour of a kNN model?

- ☐ Yes
- ☐ No

0 votes • Show Results

Vote

Pinned

Temitayo Olugbade... /teacher/
2 hours ago

Would differing scales across features affect the behaviour of a DECISION TREE model?

- ☐ Yes
- ☐ No

0 votes • Show Results

Vote

Pinned

Temitayo Olugbade... /teacher/
2 hours ago

Would differing scales across features affect the behaviour of a LINEAR REGRESSION model?

- ☐ Yes
- ☐ No

0 votes • Show Results

Vote

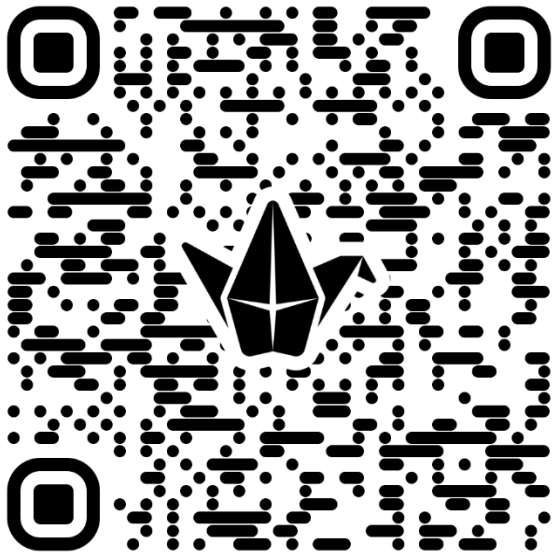
Post your notes here

Pinned

Temitayo Olugbade... /teacher/
2 months ago

Reminder:

Post your group's conclusions. Make the subject/heading the question discussed. Try to capture all of the main points from your discussion. Then, look through the main points from other groups.

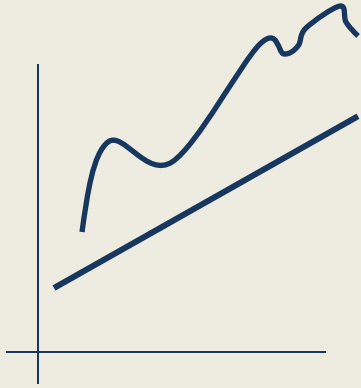


Discussion outline

- ☐ Would differing feature scales affect the behaviour of the model?
- ☐ A student-posed question
- ☐ What does this AI do?



Learning algorithms so far



Linear regression



Decision tree



kNN

Toy data (Stats)

		FEATURES				
		Humidity	No. of days of sunshine	Wind speed	Temperature	Rainfall rate
Descriptive stats	Mean	82.52	120.345	4.67	9.42	97.68
	Standard deviation	5.16	63.21	1.18	4.51	68.33
	Min	63.03	3.49	2.11	-1.62	0.28
	Max	95.93	345.34	11.82	20.37	697.13

? Class question – What do you notice about the scale of the features?

Hint – Think about how scale compares across features.

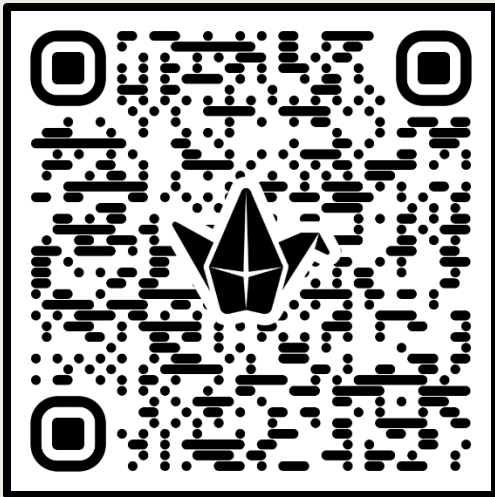
Features' scales



Student-student discussion

For each of the 3 types of models, would differing scales across features affect the behaviour of the model? *And why or why not?*
(20mins)

Hint – Think about how the models get trained.

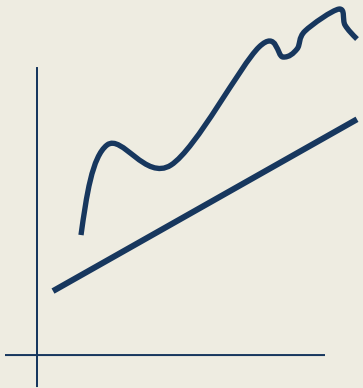


& How could one address differing feature scales, if an issue? **(5mins)**

Toy data

		<i>D</i> FEATURES						
		hurs	psl	sun	pv	sfcWind	tas	rainfall
<i>N</i> DATA INSTANCES	1	89.45	1022.71	57.49	8.81	5.88	6.23	116.35
	2	84.77	1020.19	76.77	8.03	4.38	5.47	50.66
	3	87.87	1007.45	19.61	5.65	6.20	-0.59	92.88

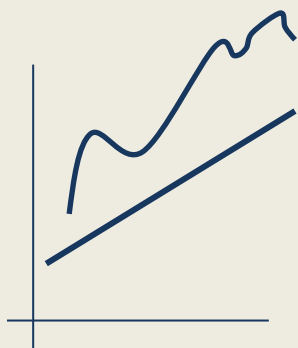
Re: ML with the basic linear model



- Get training data, i.e. $(\mathbf{x}_n, \mathbf{y}_n)$ pairs, $1 \leq n \leq N$
- Choose an error metric, e.g. mean-squared error
- Find the optimal model parameters, i.e. the best values for \mathbf{w}^* and b^* , e.g. $\mathbf{w} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}$
- Plug this in your model and apply to obtain $\hat{\mathbf{y}}$

$$f(\mathbf{x}) = \hat{\mathbf{y}} = \mathbf{x}\mathbf{w}^* + b^*$$

Re: ML with the basic linear model (2)



- Get training data, i.e. $(\mathbf{x}_n, \mathbf{y}_n)$ pairs, $1 \leq n \leq N$
- Choose an error metric
- Find the optimal model parameters, i.e. the best values for \mathbf{w}^* and b^*
- Plug into your model and apply to obtain $\hat{\mathbf{y}}$

$$f(\mathbf{x}) = \hat{\mathbf{y}} = \mathbf{x}\mathbf{w}^* + b^*$$

$$\hat{\mathbf{y}} = b + w_{hurs}hurs + w_{psl}psl + \cdots w_{rainfall}rainfall$$

Re: Decision tree algorithm



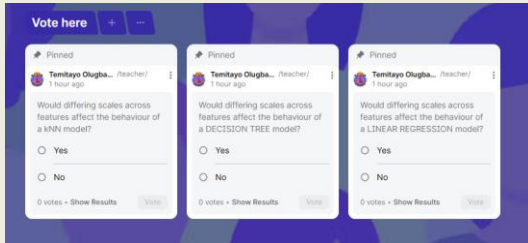
1. Given purity threshold, max number of leaf nodes, and initial region, compute the purity of the region
2. For the region, if purity \geq threshold:
 - Stop splitting
 - Make majority class in the region the leaf node
3. Else:
 - For each feature, find the best split point based on a split criterion
 - ❖ For a given feature, sort the values for that feature in the dataset
 - ❖ Compute potential split points – Midpoint between each value and the next, if different
 - Then, find the feature with the best split point for the current region
 - Split the region using that feature and its split point
 - For each region in the split, repeat from (1)

Re: How a kNN ~~training~~ & inference works



- For each given test instance, find its k nearest neighbours based on a distance metric, e.g. Euclidean distance
- Determine the class \hat{y}_i of x_i from the labels y_1, y_2, \dots, y_k of these k nearest neighbours using a voting strategy

Differing feature scales



Would differing scales across features affect the behaviour of the model?

LINEAR REGRESSION	DECISION TREE	KNN
Yes ✓	No ✗	Yes ✓

ALWAYS scale your features before training (and inference).

Feature scaling (aka normalization)

To ensure that feature dimensions x_d have similar distribution of values $\forall d, 1 \leq d \leq D$

Common feature scaling methods

- Standard scaling

scales to mean $\check{\mu}_d = 0$ and standard deviation $\check{\sigma}_d = 1, \forall d$

$$\check{x}_d^n = \frac{x_d^n - \mu_d}{\sigma_d}$$

- Min-max scaling

default min-max scales to min $\check{x}_d^{\min} = 0$ and max $\check{x}_d^{\max} = 1, \forall d$

$$\check{x}_d^n = \frac{x_d^n - x_d^{\min}}{x_d^{\max} - x_d^{\min}}$$

\check{x}_d^n, x_d^n = scaled feature, original feature at dimension d for data instance n

μ_d, σ_d = mean, standard deviation for feature at d across all $n, 1 \leq n \leq N$

x_d^{\min}, x_d^{\max} = min, max value at d across all $n, 1 \leq n \leq N$

Standard scaling with ScikitLearn

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(feats)
scaled_feats = scaler.transform(feats)
print("\n A peek at the scaled dataset features: \n"+str(scaled_feats))
```

Any questions???

Discussion outline

- ☐ Would differing feature scales affect the behaviour of the model?
- ☐ **A student-posed question**
- ☐ What does this AI do?



A student-posed question (1)



Class question

Could a dataset contain only one feature? Why (with real-world examples) or why not?

A student-posed question (2)



Student-student discussion

What if a dataset only contains one feature? How can a decision tree be used here? **(10mins)**

Hint – Use the decision tree algorithm to manually build a decision tree.

Feature	Label
6.23	1
5.47	1
-0.59	0

Re: Decision tree algorithm

1. Given purity threshold, max number of leaf nodes, and initial region, compute the purity of the region

Purity = proportion of instances of the majority label = 0.67

2. Until purity \geq threshold:

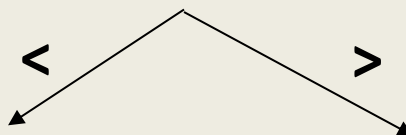
- For each feature, find best split point based on split criterion

Potential split points are 2.44, 5.85

Split point 2.44 has higher information gain

- ~~○ Then, find the feature with the best split point for the current region~~
- Split the region using that feature and its split point

Feature = 2.44




Label = 0


Label = 1

- ~~○ For each region in the split, repeat from (1)~~

initial region

Feature	Label
6.23	1
5.47	1
-0.59	0

 purity threshold=1

 maximum number of leaf nodes= ∞

Any questions???

Discussion outline

- ❑ Would differing feature scales affect the behaviour of the model?
- ❑ A student-posed question
- ❑ **What does this AI do?**



Choose TWO AIs to dissect and discuss



Temitayo Olugbade • 21d

Add and Vote for an AI to dissect!

Add AI tools (name and url) that we could dissect in a lecture session. Vote for AI listed.

 Temitayo Olugbade... /teacher/ 6 months ago

DRISTi

An AI tool for screening the eyes for diabetic retinopathy (an eye condition caused by diabetes)

<https://artelus.com/dristi-screening>

👍 0 🗳️ 0


 Temitayo Olugbade... /teacher/ 6 months ago

Cogniti

A platform for building custom AI to support teaching and learning

<https://cogniti.ai/>

👍 0 🗳️ 0


 Temitayo Olugbade... /teacher/ 6 months ago

AI-DOMP

An AI system for assessing the potential of mould growth

<https://ai-domp.com/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 6 months ago

Charismatic

An AI for story generation

<https://www.charismatic.ai/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 6 months ago

Octave

An emotion-aware voice AI

<https://www.hume.ai/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 6 months ago

Copilot

A virtual assistant for text and image processing/generation tasks

<https://copilot.microsoft.com/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 2 months ago

IONOS

AI for website building, domain registration, and other (e.g. content generation, customer support)

<https://www.ionos.co.uk/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 2 months ago

Affiniti AI

An AI mental health co-therapist

<https://www.affiniti.ai/>

👍 0 🗳️ 0

 Temitayo Olugbade... /teacher/ 2 months ago

Fleet Driver Monitoring System

An AI for monitoring driver behaviour

<https://www.smarteve.se/solutions/automotive/fleet-and-aftermarket/>

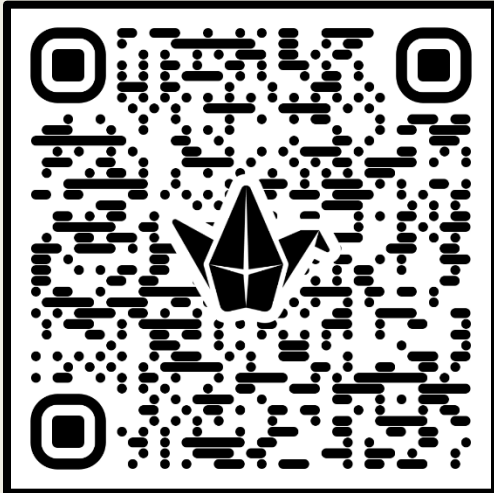
👍 0 🗳️ 0

Dissect the AI



Student-student discussion

- What exactly does it do? How is it used? **(10mins)**
- What specific AI models (i.e. tasks/capabilities) are or could be involved? **(10mins)**
- What do these models do? What are or could be their input data? What are or could be their output (label)? **(10mins)**



Any questions???