Week 3 discussion

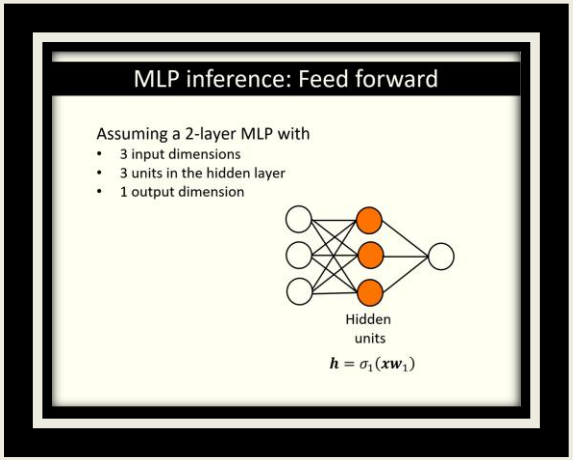
MACHINE LEARNING

Dr. Temitayo Olugbade

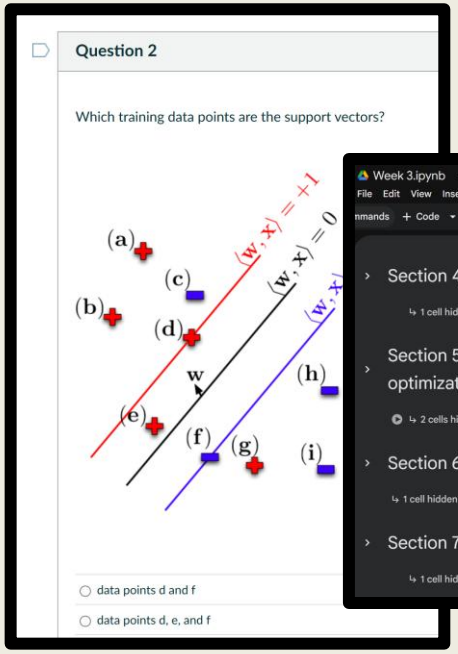
UNIVERSITY
OF SUSSEX

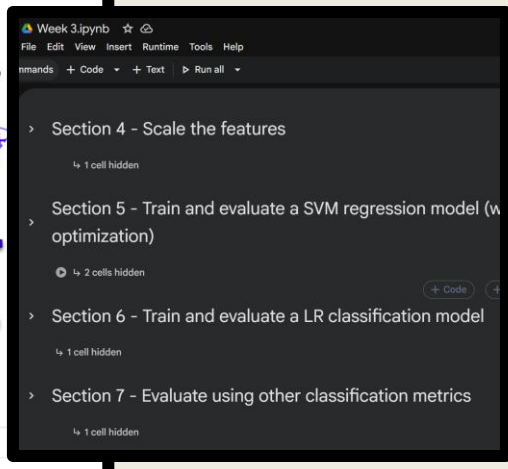
Student To Do

- **Watch Week 4 mini-videos ahead of next lecture**



- **Get to Week 3 code notebook & ungraded quiz when you can soon**





Week 3 mini-video content

This week, you've been looking at:

- Support vector machines
- Probabilistic models

Student questions for Week 3

4 Replies

Frequently Asked Questions (FAQs)

This discussion forum will be used to provide answers to FAQs from students.

You can add questions to Student questions - Machine learning.

Reply

TO

Temitayo Olugbade Teacher

4 Feb 16:23 | Last edited 6 Feb 16:35

View History

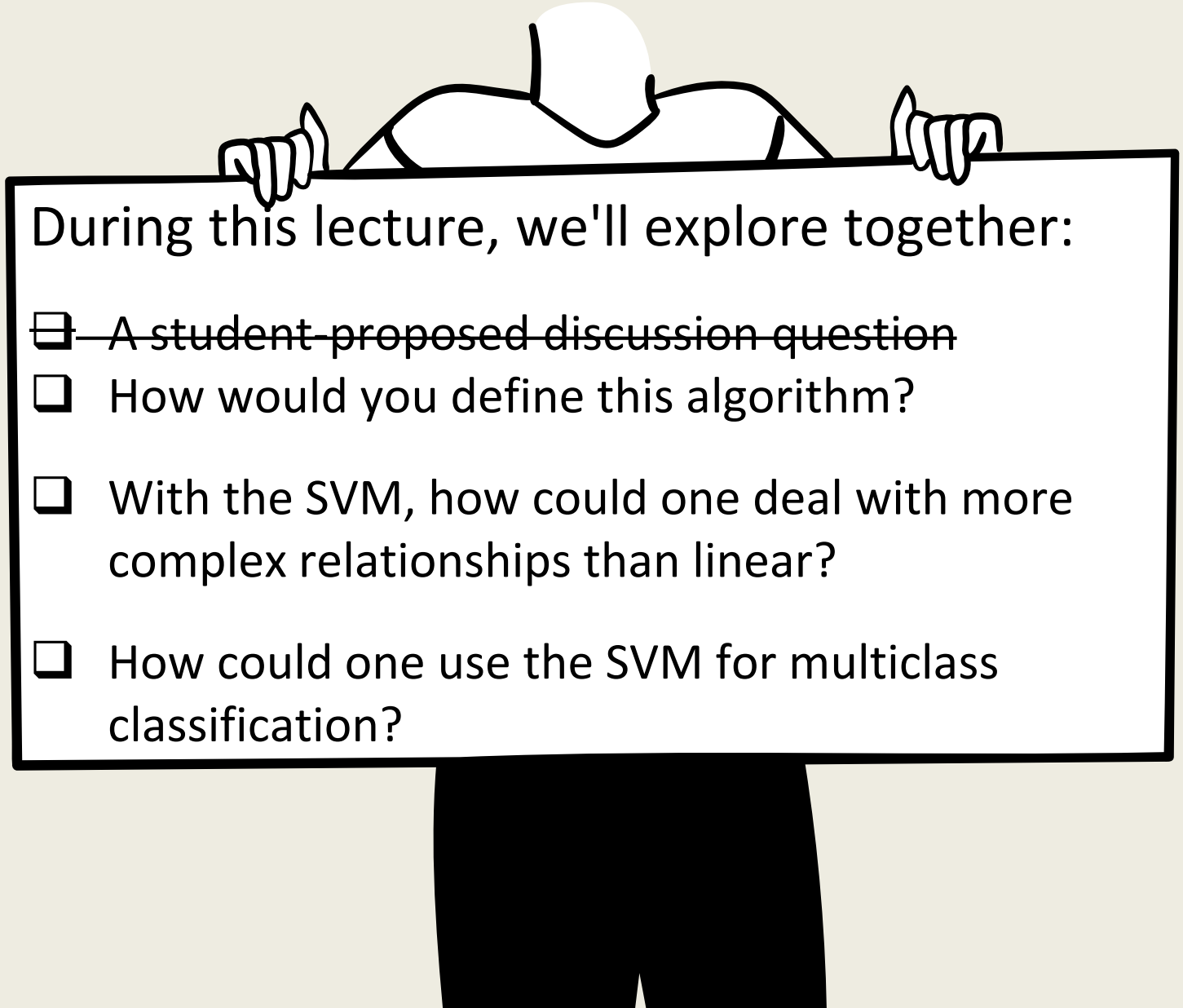
STUDENT QUESTION

Could you go through an example of using k-d trees to reduce the complexity of kNN?

TUTOR RESPONSE (UPDATED)

A good place to get started on digging more into kd-trees is the original introduction of the method:

Learning outcomes



During this lecture, we'll explore together:

- ~~A student-proposed discussion question~~
- How would you define this algorithm?
- With the SVM, how could one deal with more complex relationships than linear?
- How could one use the SVM for multiclass classification?

Discussion outline

- ❑ In your own words… **(30mins)**
- ❑ What are the assumptions of the SVM? What if they aren't met? **(30mins)**
- ❑ How could one use the SVM for multiclass classification? **(30mins)**

Discuss & Note down summary

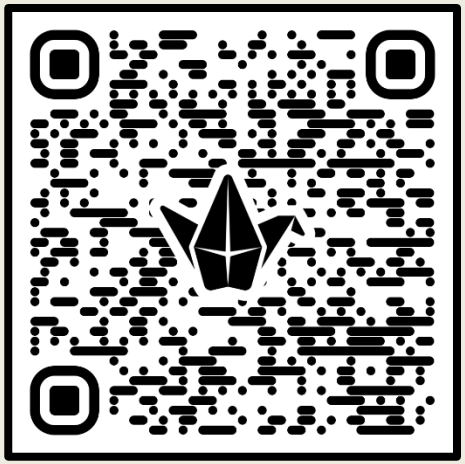
In your own words…

- **In you own words…**
- What are the assumptions of the SVM? What if they aren't met?
- How could one use the SVM for multiclass classification?

Naïve Bayes model

Student-student discussion – **In your own words**, how would you define a **Naïve Bayes model**? *(15mins)*

- Take 2 mins to think about how you would define it, WITHOUT searching lecture slides, textbook, Google, ChatGPT, etc.
- In groups of 3, take turns in
  - ✓ 1 person sharing their own definition.
  - ✓ The other two asking 1 question or adding information (1 sentence).
- Check your definitions against the lecture slide's etc.
- Type up the group's agreed definition.



Logistic regression learning algorithm

Student-student discussion – **In your own words**, how would you define a **logistic regression learning algorithm**? *(15mins)*

- Take 2 mins to think about how you would define it, WITHOUT searching lecture slides, textbook, Google, ChatGPT, etc.
- In groups of 3, take turns in
  - ✓ 1 person shares their own definition.
  - ✓ The other two asking 1 question or adding information (1 sentence).
- Check your definitions against the lecture slide's etc.
- Type up the group's agreed definition.

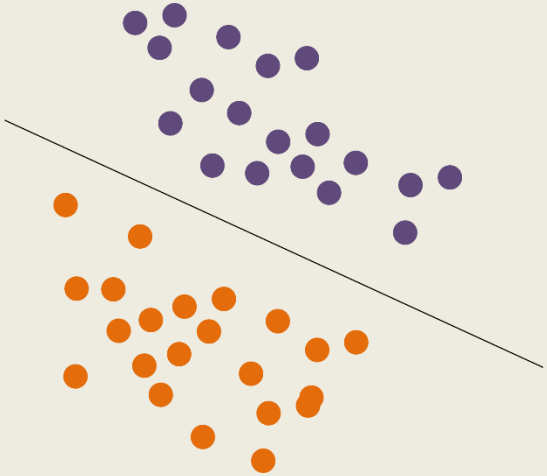
Assumptions of the SVM

- In you own words…
- **What are the assumptions of the SVM? What if they aren't met?**
- How could one use the SVM for multiclass classification?

What are the SVM assumptions?

Class question – What are the SVM assumptions?

Hint – Think about SVM definition – The SVM is the optimal hyperplane (a) at the maximum margin from two classes; (b) while still correctly classifying them



Assumption #1

The SVM is the optimal hyperplane

(a) at the maximum margin from **two classes**;

(b) while still correctly classifying them.

Assumption #2 – Perfect linear separability

Linearly separable

Somewhat linearly separable

Not linearly separable

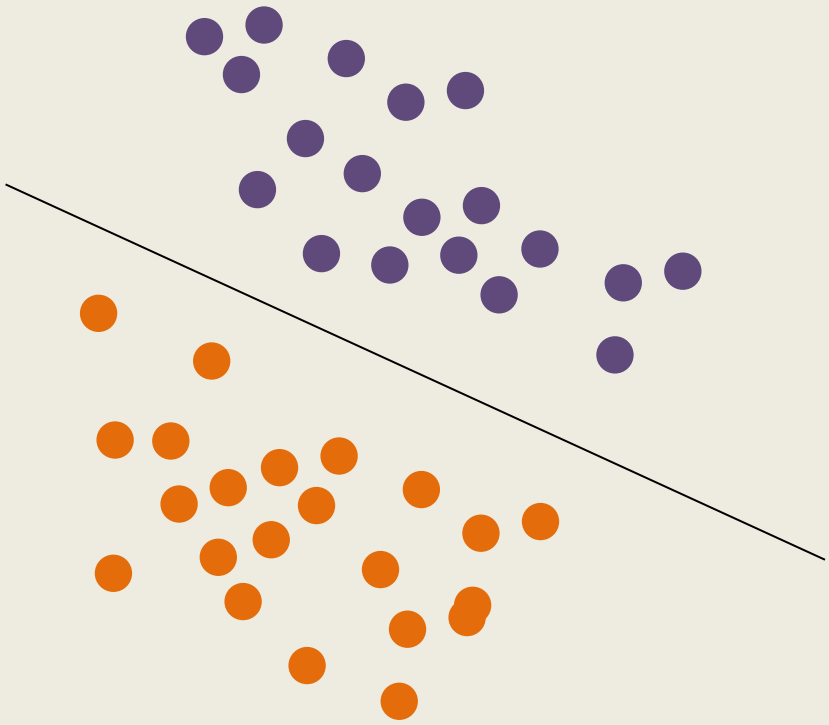
What if the assumptions are not met?

Student-student discussion – How could you use the SVM with non-linear separable classes? **(20mins)**

Hint

- What could you do to the data?
- What could you do to the algorithm?

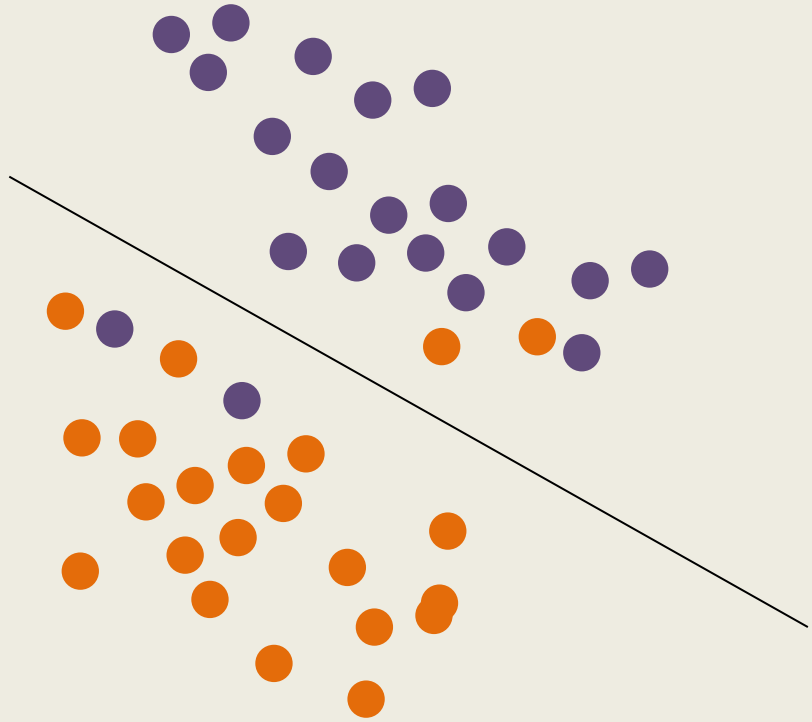
Ordinary SVM has a hard margin



$$y_n(\boldsymbol{x_n w}) \geq 1 \quad \forall n$$

Hard margin, i.e. all points in class -1 lie on one side of the decision boundary (hyperplane) & same for class +1

SVM with soft margin



$$y_n(\boldsymbol{x_n w}) \geq 1 - \xi_n \quad \forall n$$

Soft margin, i.e. allow for some errors

Soft margin as box constraint (L1 error)

The optimal hyperplane with a soft margin is equivalent to minimizing

$$\frac{\|\boldsymbol{w}\|^2}{2} + C\sum_{n=1}^{N}\varepsilon_n$$

such that $y_n(\boldsymbol{x_n w}) \geq 1 - \varepsilon_n \quad \forall n$

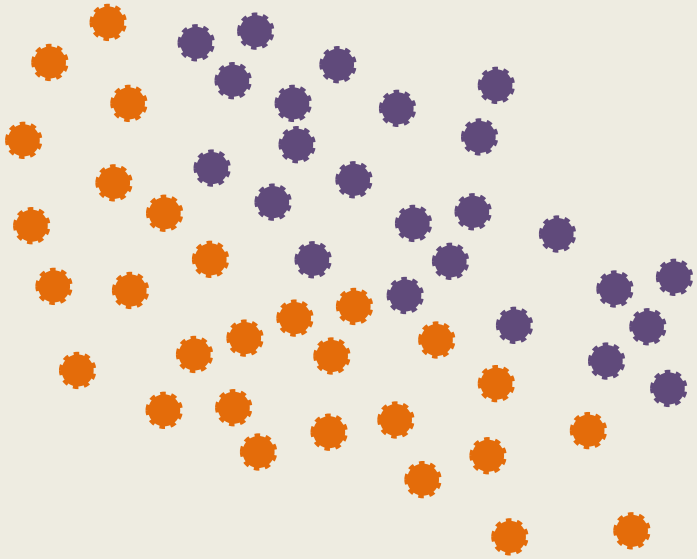
where $\varepsilon_n \geq 0$

C = box constraint

SVM with kernel trick

- Pull apart data instance pairs $(\boldsymbol{x}_n, \boldsymbol{x}_m)$ of different classes *-1* and *+1* into a new dimensional space $K(\boldsymbol{x}_n, \boldsymbol{x}_m)$
- Such that classes become separable in the new space

Original feature space



Feature space based on kernel transformation

Using a kernel

- Rewriting the dual formulation with a kernel function:

$$g(\beta) = \sum_{n=1}^{N} \beta_n - \frac{1}{2} \sum_{n,m=1}^{N} \beta_n \beta_m y_n y_m K(\boldsymbol{x}_n \boldsymbol{x}_m)$$

where $K(\boldsymbol{x}_n, \boldsymbol{x}_m)$ = kernel function

- The default kernel $K(\boldsymbol{x}_n, \boldsymbol{x}_m)$ is the linear kernel (equivalent to none), i.e. $K(\boldsymbol{x}_n, \boldsymbol{x}_m) = \boldsymbol{x}_n \boldsymbol{x}_m$

Common non-linear kernels

- Gaussian (or Radial Basis Function) kernel of standard deviation σ:

$$K(\boldsymbol{x}_n, \boldsymbol{x}_m) = e^{-\frac{(\boldsymbol{x}_n - \boldsymbol{x}_m)^2}{2\sigma^2}}$$

- Polynomial kernel of degree d:

$$K(\boldsymbol{x}_n, \boldsymbol{x}_m) = (\boldsymbol{x}_n \cdot \boldsymbol{x}_m + 1)^d$$

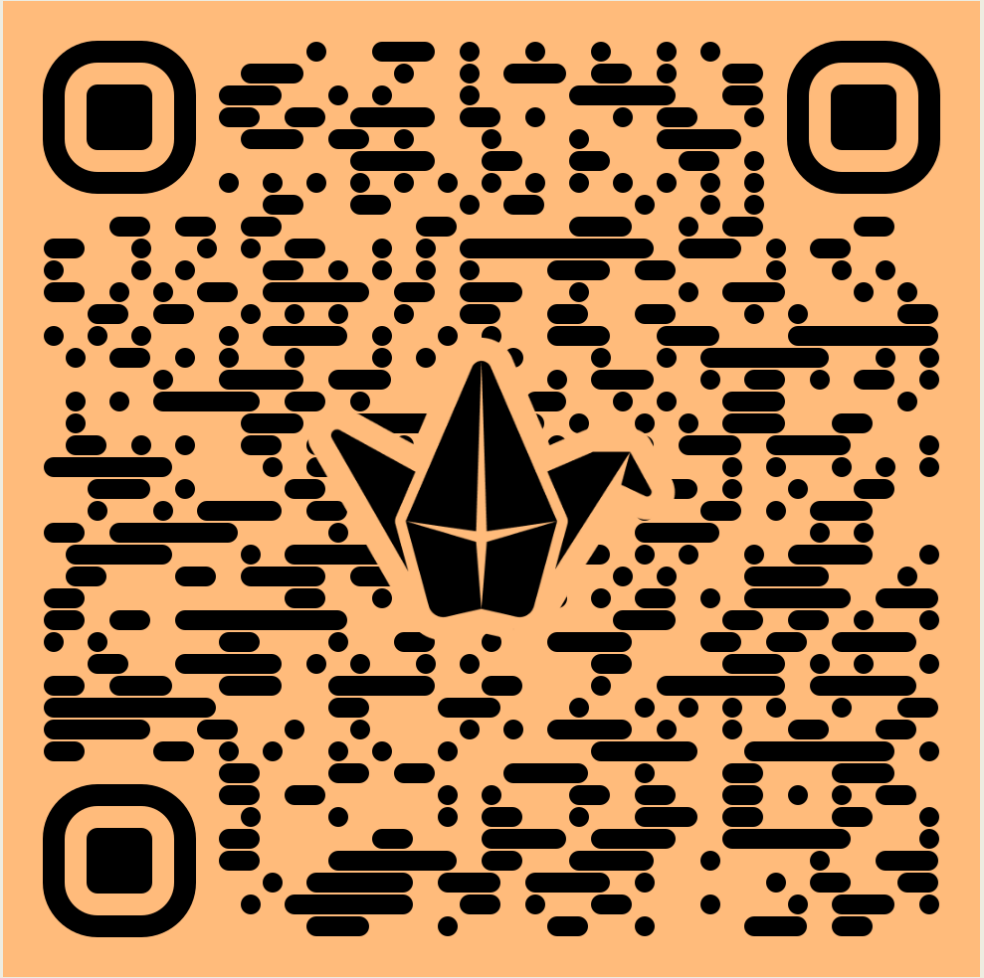
- Sigmoid / hyperbolic tangent kernel of slope γ and intercept c_0:

$$K(\boldsymbol{x}_n, \boldsymbol{x}_m) = \tan(\gamma \boldsymbol{x}_n \cdot \boldsymbol{x}_m + c_0)$$

Coursework

- In you own words…
- What are the assumptions of the SVM? What if they aren't met?
- **How could one use the SVM for multiclass classification?**

Propose a question for next week!



Assumption #1

The SVM is the optimal hyperplane

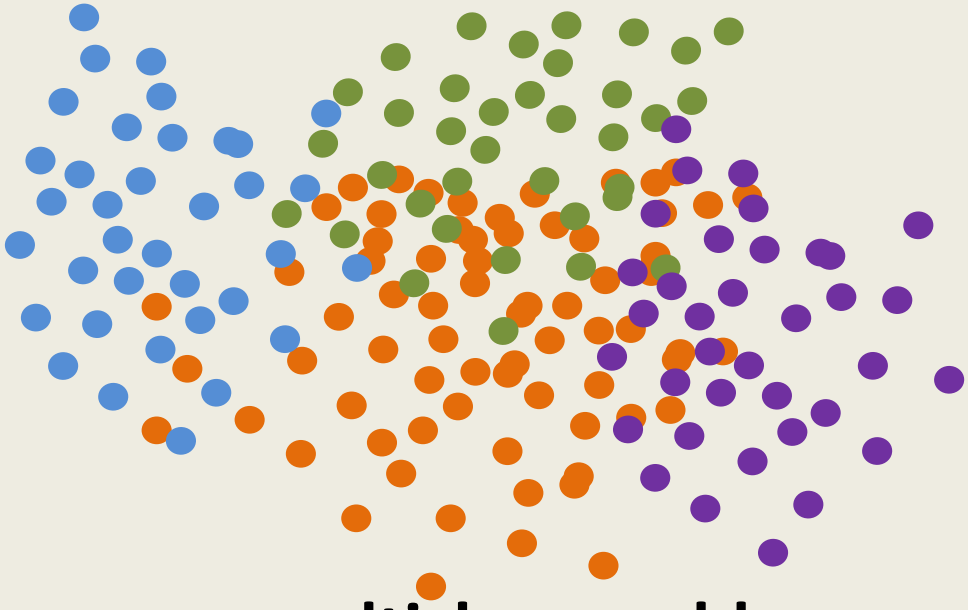
(a) at the maximum margin from **two classes**;

(b) while still correctly classifying them.

SVM for multiclass classification

Student-student discussion – How could you use the SVM for multiclass classification instead of just binary?

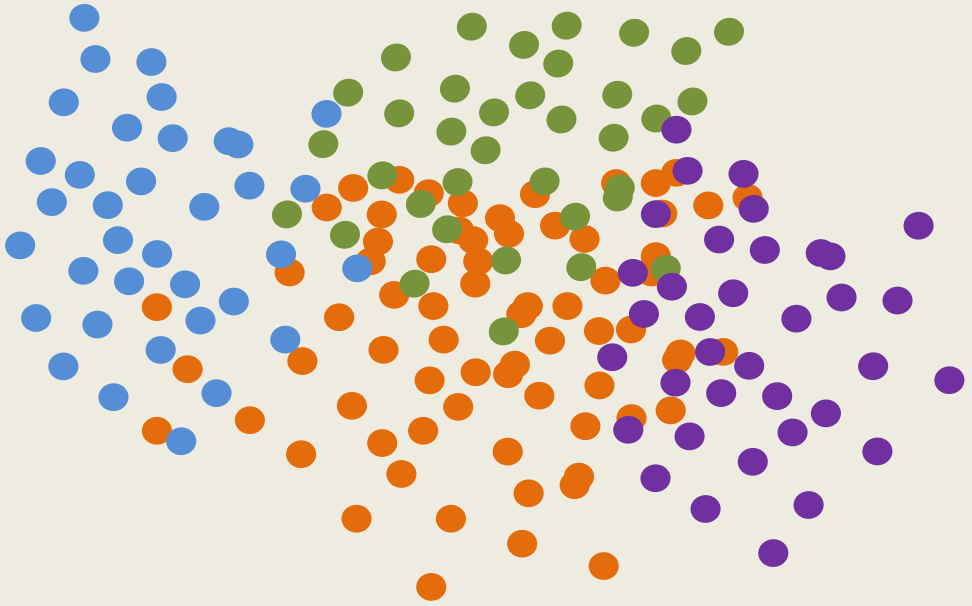
binary problem



multiclass problem

Multiclass problems

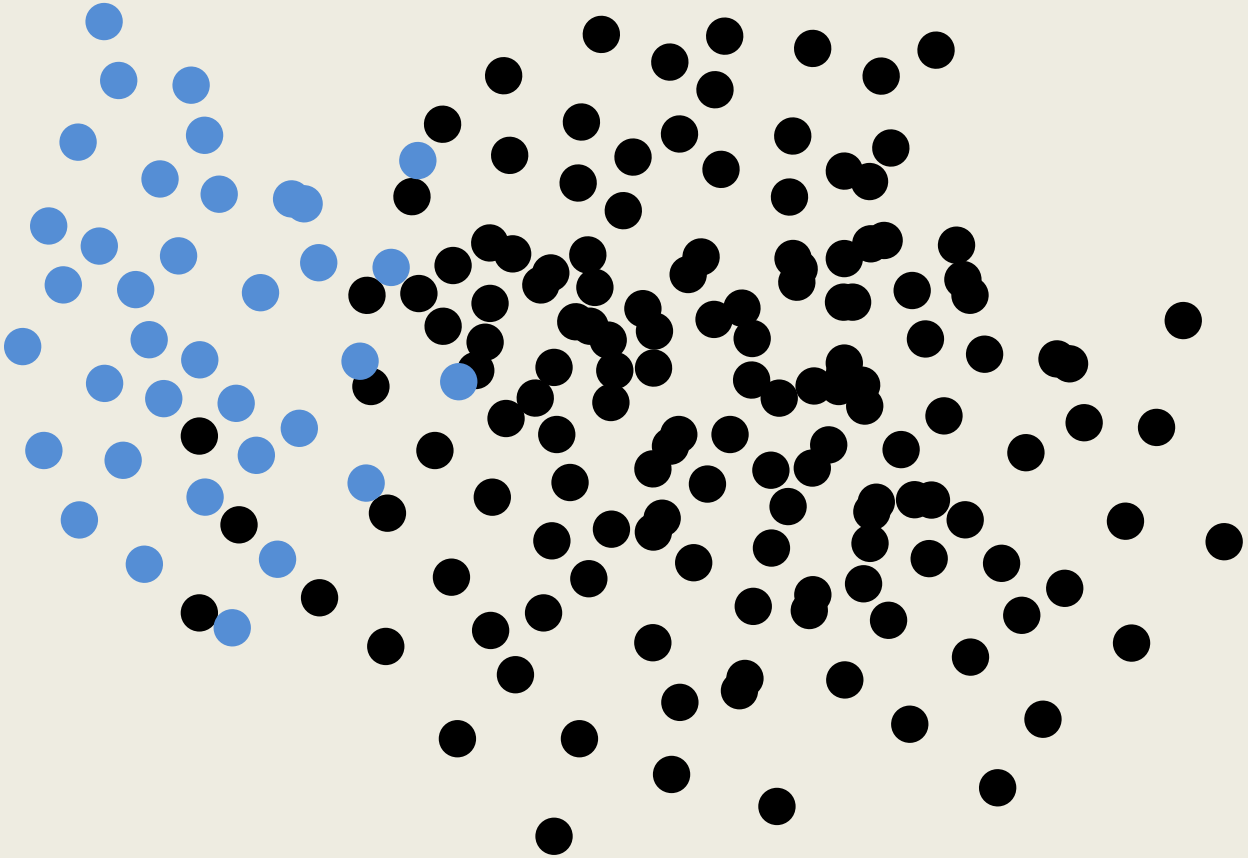
SVM is designed for binary classification



For use of SVM on multiclass problems, **one-vs-all** or **one-vs-one** is employed.

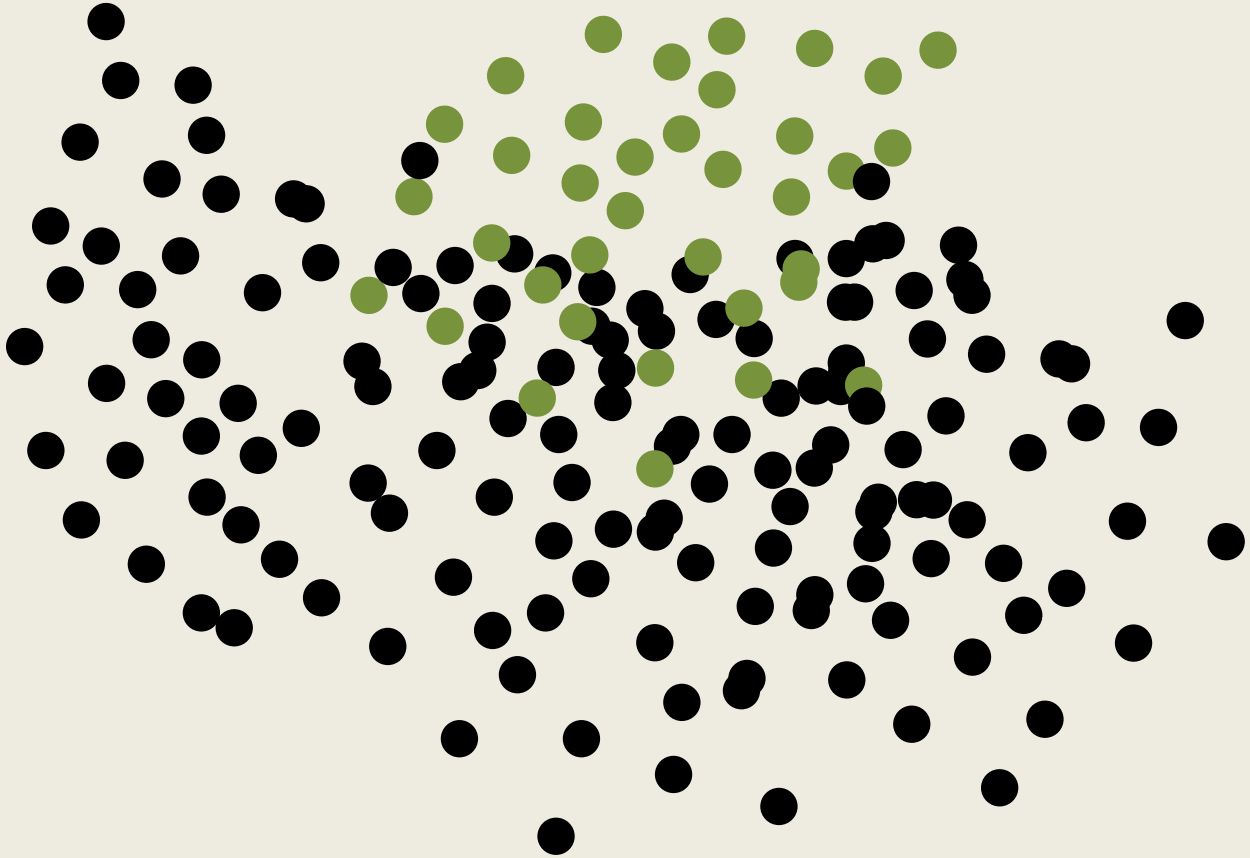
One-vs-All

	Blue	Green	Orange	Purple
Blue v All	+1	-1	-1	-1



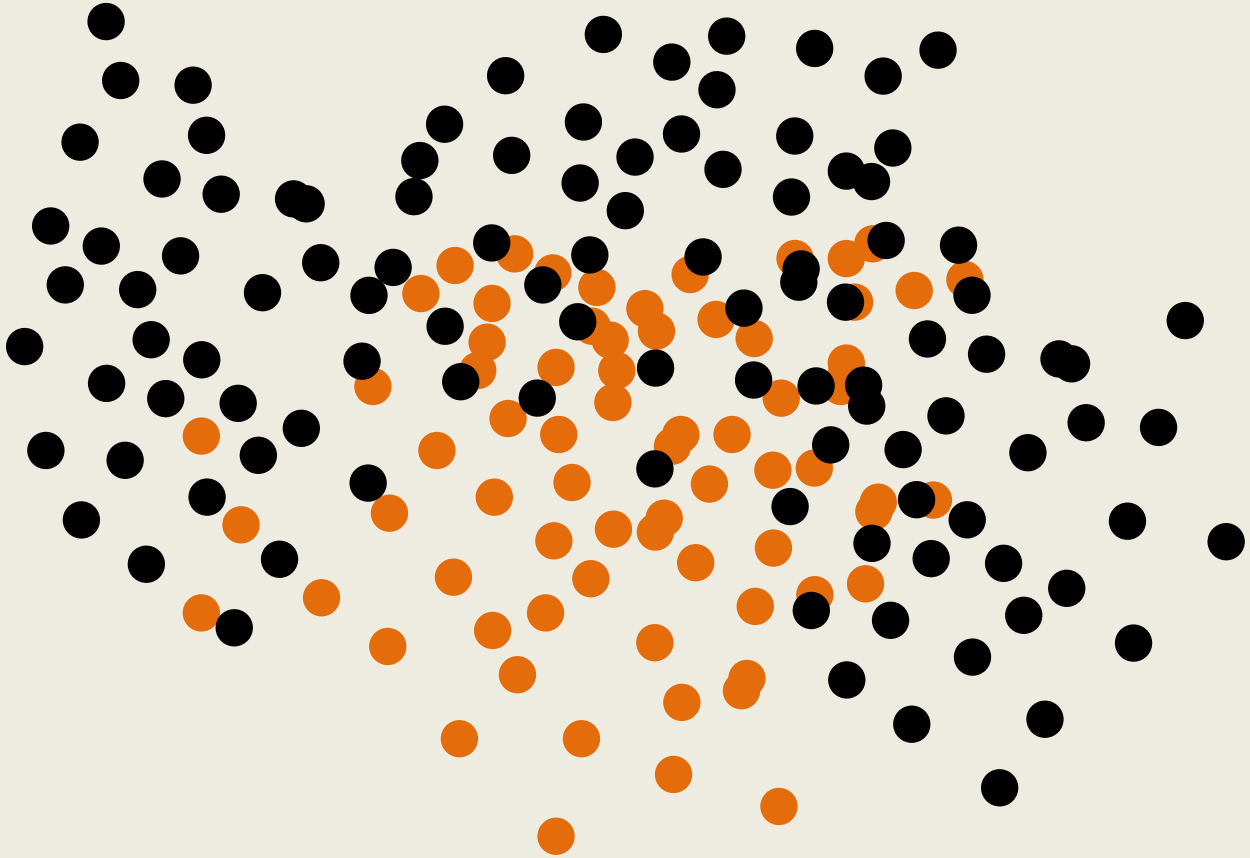
One-vs-All (2)

	Blue	Green	Orange	Purple
Green v All	-1	+1	-1	-1



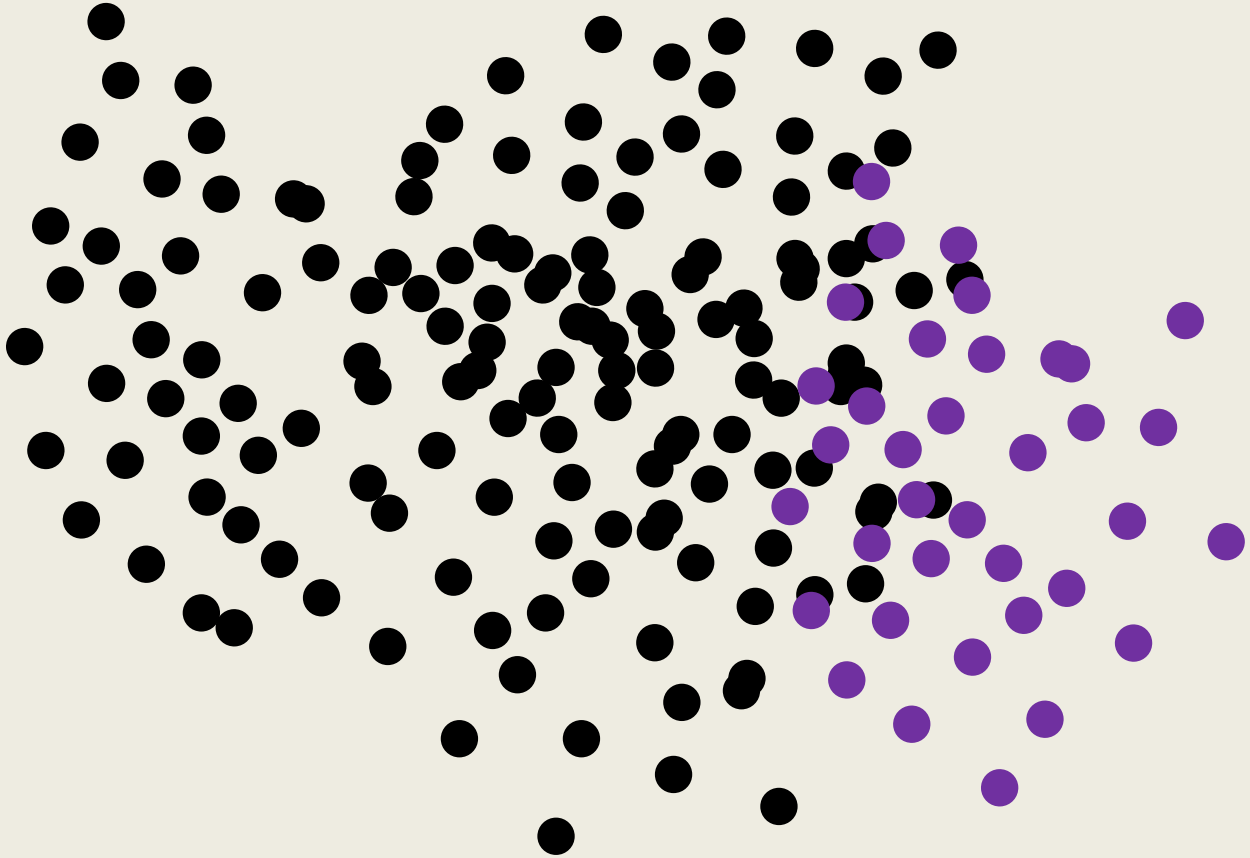
One-vs-All (3)

	Blue	Green	Orange	Purple
Orange v All	-1	-1	+1	-1

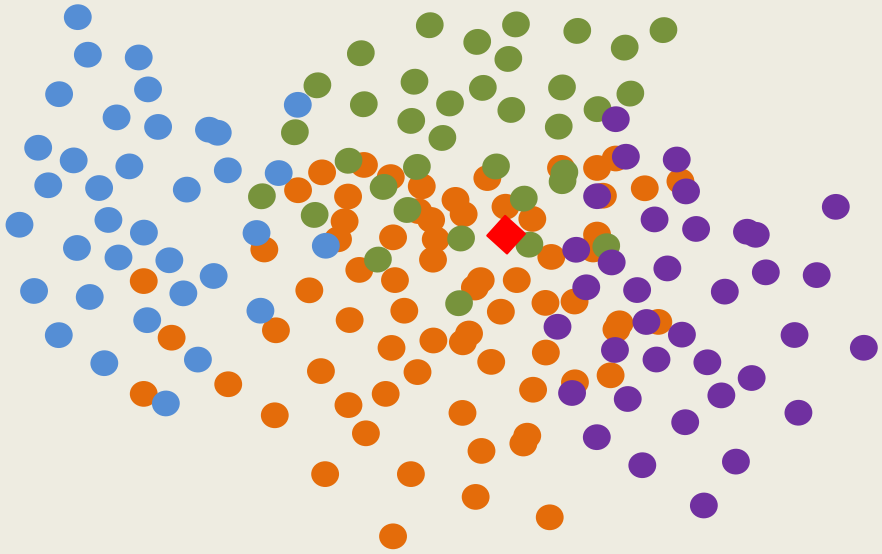


One-vs-All (4)

	Blue	Green	Orange	Purple
Purple v All	-1	-1	-1	+1



Model inference for a new data instance



New data instance (input)

Blue v All SVM model

-1 New data instance (input)

Green v All SVM model

+1

Model inference → Class ‘Green’