Segmentation in Feature Space

Dr. Ruisheng Su, Assistant Professor

Department of Biomedical Engineering, Medical Image Analysis group

IMAG/e TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Overview & course schedule

Modules	Date	Topic
Registration	April 24 (Thursday)	Course introduction, geometrical transformations
	April 28 (Monday)	Point-based image registration
	May 1 (Thursday)	Intensity-based image registration
Segmentation	May 8 (Thursday)	Introduction and evaluation metrics for image segmentation
	May 15 (Thursday)	Segmentation in feature space
	May 19 (Monday)	Segmentation using graph-cuts
	May 22 (Thursday)	Statistical shape models
Deep learning for MIA	May 26 (Monday)	Convolutional neural networks
	June 2 (Monday)	Deep learning applications (registration)
	June 5 (Thursday)	Guest lecture by **Danny Ruijters** (principal scientist @ Philips, full professor @ TU/e)
	June 10 (Tuesday)	Deep learning applications (segmentation)
	June 12 (Thursday)	Unsupervised deep learning for medical image analysis

Outline

- Recap of previous lecture
- Segmentation by clustering
- Segmentation by classifiers
- Generalization and overfitting

Outline

- Recap of previous lecture
- Segmentation by clustering
- Segmentation by classifiers
- Generalization and overfitting

Recap of learning objectives (previous lecture)

The student can:

- Describe what problems arise with using a **threshold-based** segmentation method
- Calculate how many parameters are needed for **K Gaussians** in **p dimensions**
- Recognize **covariance** matrices and identify their properties

- Select a suitable evaluation **metric** for a given medical image analysis task
- calculate the **accuracy**, **Dice score** and **Hausdorff distance**, given an image segmentation
- interpret the results and assess the quality of the validation methods used in medical image analysis research papers

Recap

Quiz: You fit a Gaussian mixture model (GMM) with K components to p-dimensional MRI patch data, using full covariance matrices.

How many parameters must you estimate?

A. $K \times (1 + p + p^2)$

B. $K \times \left(p + \frac{p(p+1)}{2}\right) + (K - 1)$

C. $K \times \left(\frac{p(p+1)}{2}\right)$

D. $K \times p \times K$

Recap

Quiz: Which of the following statement is **true** about covariance matrix?

A. It must be diagonal with equal entries.

B. It must be symmetric, but not necessarily be diagonal.

C. It must have non-negative determinant.

D. It may have negative values on the diagonal.

Recap

Quiz: Why might accuracy be misleading for evaluating segmentation of small objects (e.g., tumors, lesions)?

A. Accuracy always overestimates performance because it ignores the model's confidence.

B. Accuracy gives equal importance to background and foreground, so a model can achieve high accuracy by predicting mostly background.

C. Accuracy is too strict and penalizes every small boundary error harshly.

D. Accuracy requires ground truth contours, which are hard to obtain for small objects.

Recap

Quiz: You need to quantify the overlap between your segmented tumor mask and the radiologist's annotation.

Which metric excels at measuring region overlap?

A. Accuracy
B. Hausdorff distance
C. Dice score

Recap

Quiz: You need to evaluate a small lesion segmentation where missing even tiny parts is critical.

Which metric should you prioritize?

A. Accuracy
B. Hausdorff distance
C. Dice score

Intended learning outcomes

The student can:

- Apply **K-Means** by hand on a small dataset.
- Apply the nearest mean, nearest neighbor classifier or similar simple **classifiers** by hand on a small dataset
- Describe **properties** of simple classifiers (type of boundary, sensitivity to scaling, sensitivity to irrelevant features)
- Explain the **effect of classifier** properties on the result of a segmentation.

- Explain what is meant by **generalization and overfitting**
- **Diagnose** generalization/overfitting given scatterplot of dataset, decision boudnary and/or classifier results (error plots)
- **Compare** classifiers based on their complexity / ability to overfit, understand how choice of parameters influences complexity
- Name **reasons overfitting** can occur / be able to suggest strategies to reduce overfitting

Outline

- Recap of previous lecture
- Segmentation by clustering
- Segmentation by classifiers
- Generalization and overfitting

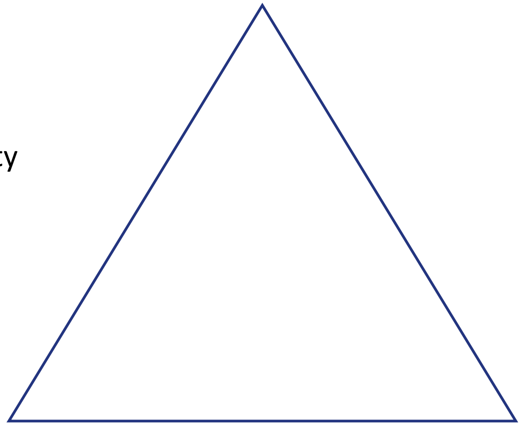
Image segmentation

Assumptions about the data

- Pixels of the same tissue type cluster in feature space
- Object boundary is smooth

Model

- Thresholding
- Clustering in feature space
- Region growing
- Graph-based models
- Learning-based models
- … …



Considerations

- Data imbalance, small structures
- Trade-off between sensitivity and specificity
- Boundary precision (worst-case errors)
- Annotation uncertainty
- Clinical perspective

Metrics

- Accuracy
- Dice coefficient
- Hausdorff distance
- … …

Optimization

- expectation-maximization for a Gaussian mixture
- k-means updates
- gradient descent

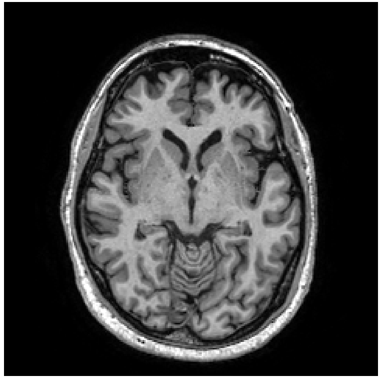
Further reading: https://medium.com/%40devnag/machine-learning-representation-evaluation-optimization-fc7b26b38fdb

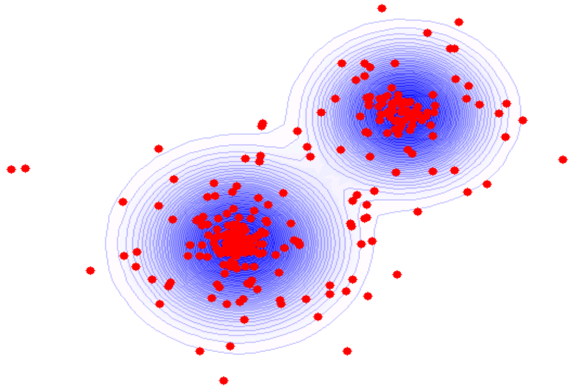
Segmentation in feature space

General model for pixel-based segmentation:

- Each pixel is a point in feature space $\underbrace{[\,I(x,y),\ \text{gradient}(x,y),\ \text{texture}(x,y)\,]}_{\text{feature vector for that pixel}}$

- Similar pixels (e.g. white matter) are close to each other in feature space

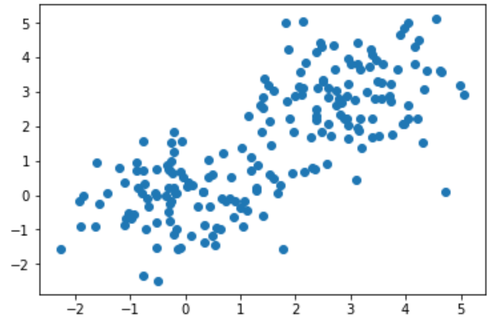




K-Means clustering

Model

→ *M* pixels are grouped into *K* clusters, pixel belongs to the cluster closest to it



Optimization

→ find *K* clusters with the “best fit”

- How many parameters do we optimize over?
- K is a *hyperparameter*, it is set outside of the algorithm

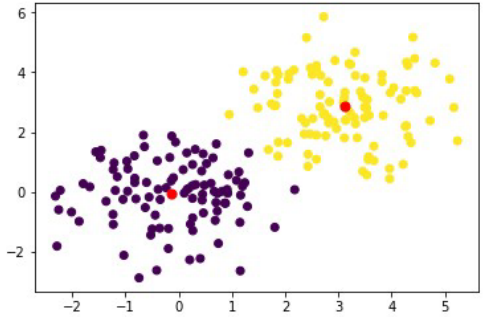
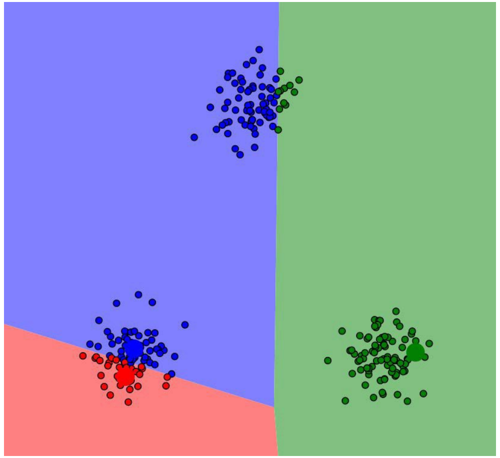


Image source: courtesy of V. Cheplygina

IMAG/e TU/e

K-Means clustering

Evaluation: How to evaluate "best fit"?



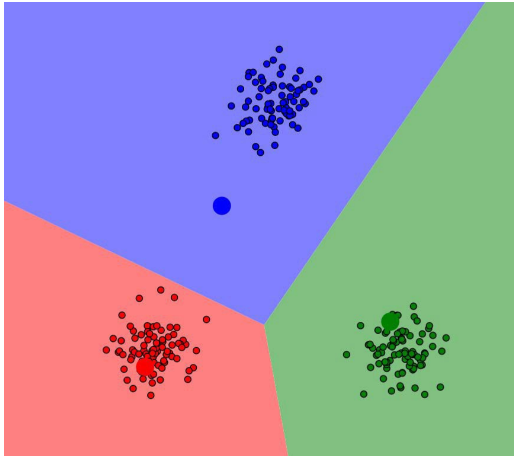


Image source: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/

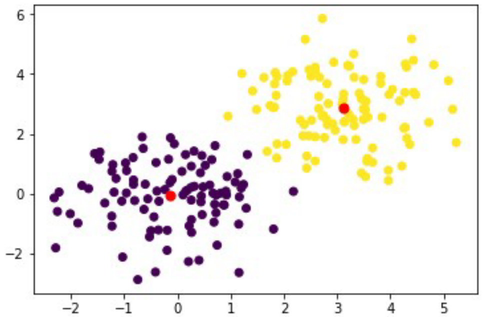
K-Means clustering

Evaluation

→ Average squared Euclidean distance between each point and the closest cluster

$$J(W) = \frac{1}{N}\sum_i \|min_k(W_k - X_i)\|_2^2$$

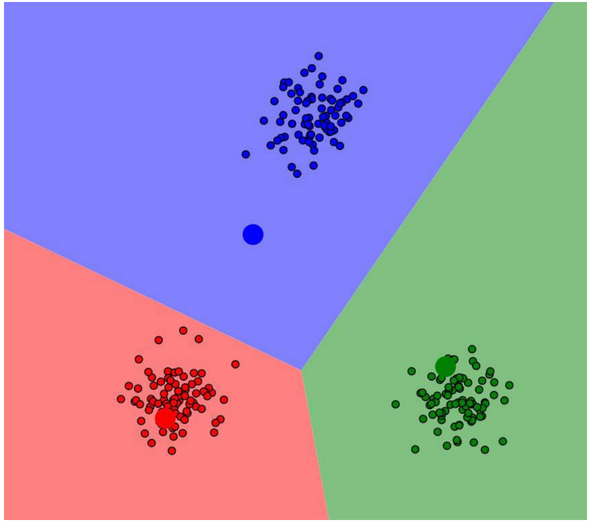
X are the points, W are the cluster centers



K-Means clustering

Optimization

- How to go through solutions so that the solution improves?
- Given a solution, what should we do to reduce the error?



K-Means clustering

Optimization

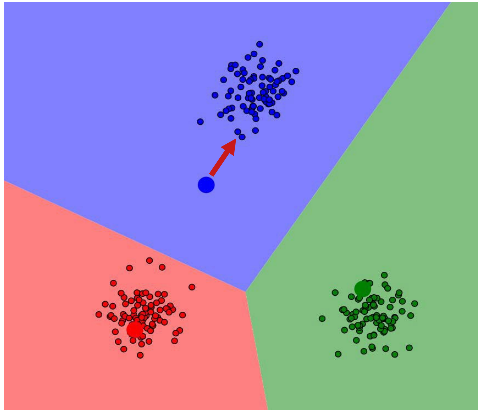
- Gradient descent, set derivative to zero

- c(X) is the cluster index of the center closest to X,
  N_k is the number of points in k-th cluster

For the *k*-th cluster:

$$\frac{\partial}{\partial W_k} J = - \frac{1}{N_k} \sum\nolimits_{c(X)==k} \| (X - W_k) \|_2^2$$

It is a vector in the direction of the new cluster center!



K-Means clustering

overall algorithm

Initialize K cluster centers

Until some stopping criterion:

- Assign all samples to nearest cluster center
- For each cluster center:
  - Calculate derivative
  - Update cluster centers

K-Means demo in the practicals

Source: https://www.youtube.com/watch?v=5I3Ei69I40s

K-Means clustering

K-means hyperparameters:

- **K** – how many cluster centers
- **Initialization** – how to select cluster centers the first time, for example “random”
- **Stopping criterion** – how to decide if to continue, for example “after 100 iterations”

K-Means clustering

Interactive demo:

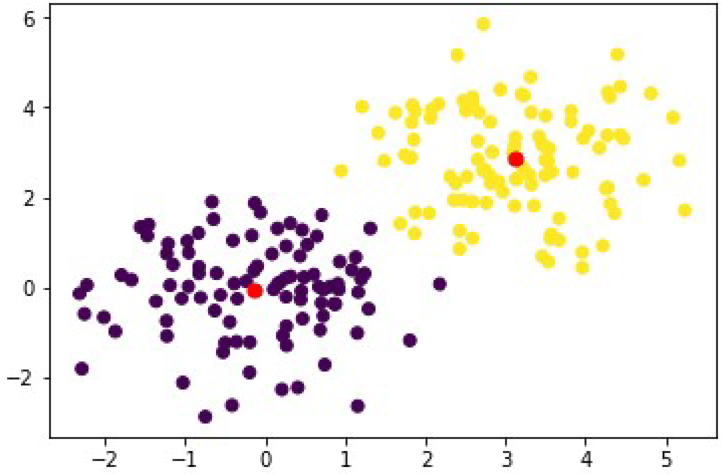
https://www.naftaliharris.com/blog/visualizing-k-means-clustering/

Self study quiz: what types of data are easy or difficult for k-Means?

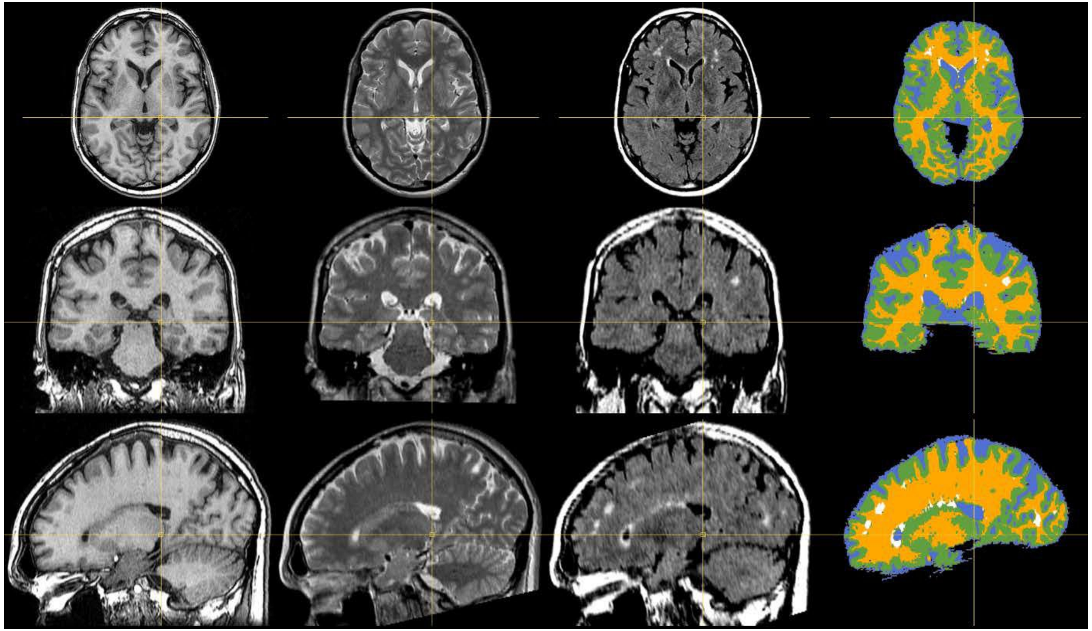
K-Means clustering

Output: which pixel belongs to which cluster.

Q:How can we use this for tissue segmentation?

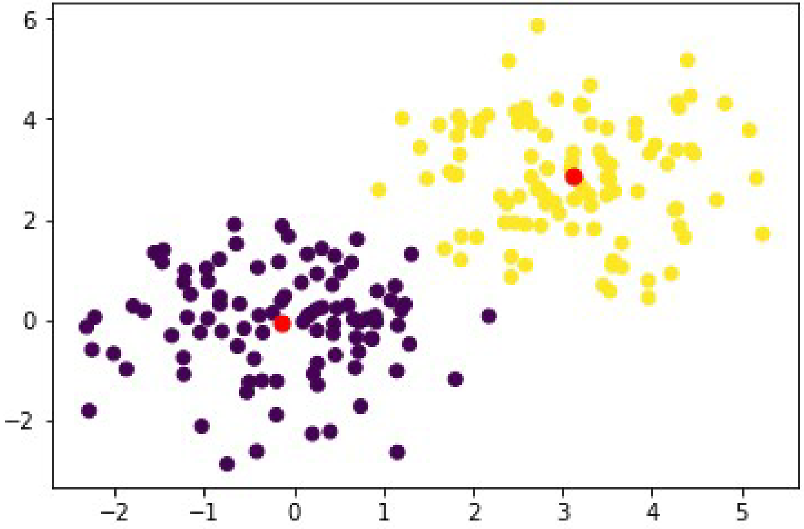


- **Step 1: tissue segmentation**



K-Means clustering

After clustering, use prior knowledge to assign clusters to classes,
e.g. brain has higher intensity than background



intensity

Outline

- Recap of previous lecture
- Segmentation by clustering
- Segmentation by classifiers
- Generalization and overfitting

Segmentation by classifiers

- Nearest mean classifier
- Nearest neighbour classifier
- Histogram-based classifier
- Decision tree classifier

Nearest mean classifier

What if we already have data that has been segmented?

- E.g. features of white matter and grey matter from one subject
- What class is this green voxel from a new subject?

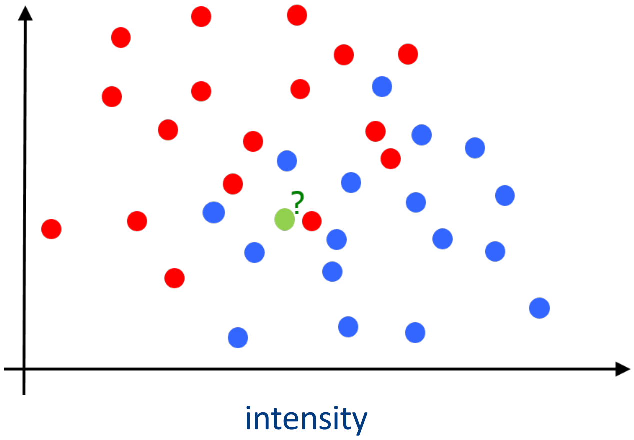


Image source: courtesy of V. Cheplygina

Nearest mean classifier

It is more similar to the average grey matter voxel

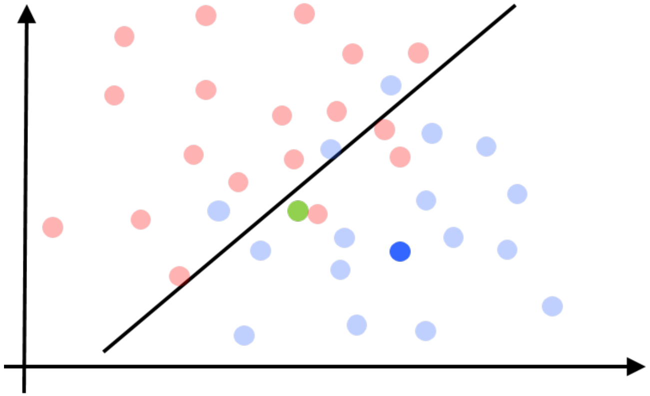


Image source: courtesy of V. Cheplygina

Nearest mean classifier

- Dataset with 5 training points in 3 dimensions
- How does the nearest mean classifier label the test point?

Class 1:

Point 1:	[3	4	4]
Point 2:	[5	5	2]
Point 3:	[1	6	3]

Class 2:

Point 4:	[2	2	4]
Point 5:	[2	4	6]

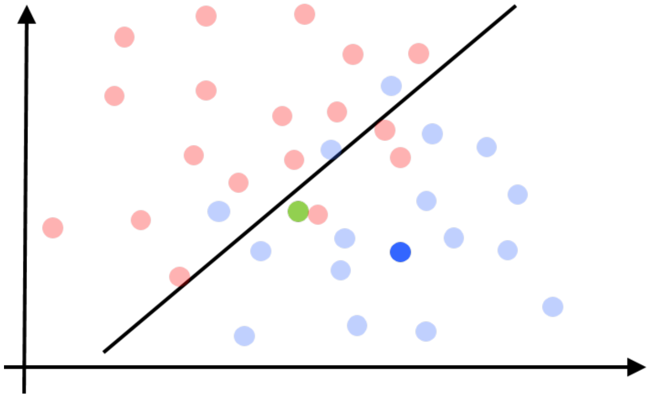
Test point:

[5 3 4]

Class 1

Nearest mean classifier

- Model each class **by its mean**
- For a test point, calculate its distances to the class means, find class with the smallest distance
- Linear boundary

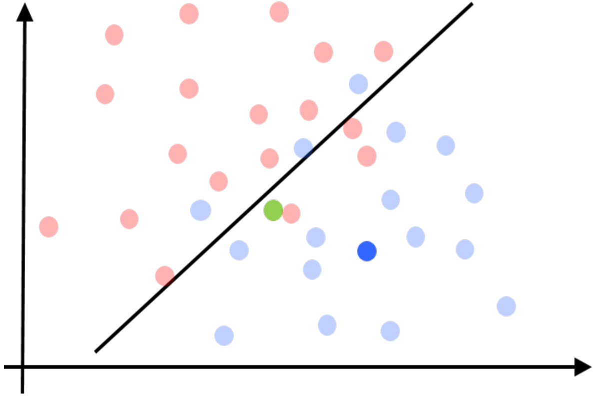


Nearest mean classifier

Nearest mean classifier is sensitive to scaling

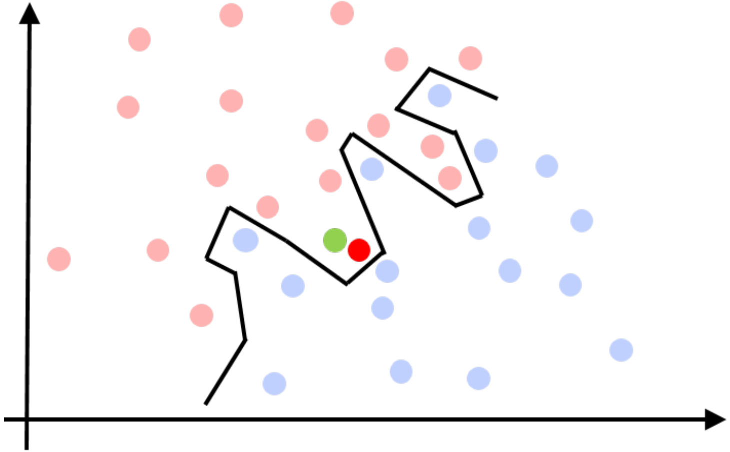
- E.g. each point is a subject, height in centimeters on x axis and weight in kilograms on y axis

- Imagine instead plotting height in meters



Nearest neighbor classifier

It is very similar to a white matter voxel



Nearest neighbor classifier

How does the nearest neighbor classifier label the test point?

Class 1:

[3	4	4]
[5	5	2]
[1	6	3]

Test point:

[5	3	4]

Class 2:

[2	2	4]
[2	4	6]

Class 1

Nearest neighbor classifier

- Model each class by **all of its training points (**“lazy learning”)
- For a test point, calculate its distances to all training points, find class with the smallest distance
- Non-linear boundary
- Sensitive to feature scaling

Nearest neighbor vs nearest mean

	Nearest mean classifier	Nearest neighbor classifier
Model	Each class is represented by its mean	Each class is represented by all of its points
Optimization (training)	Calculate class means	Lazy learning

Nearest neighbor classifier

- Local changes in training data have large influence on the boundary
- If test data = training data, error = 0

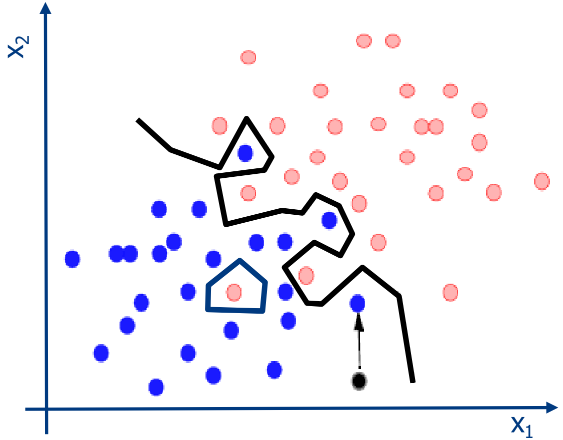


Image source: David Tax

K-nearest neighbor classifier

- To reduce variability, look at majority of k nearest neighbors
- Use odd k to avoid ties
- Less sensitive to local changes / boundary is more smooth

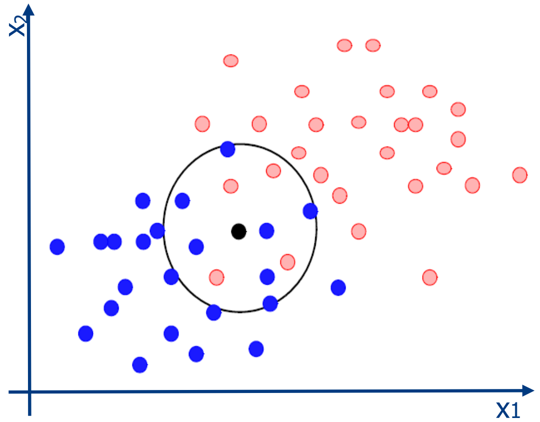


Image source: David Tax

K-nearest neighbor classifier

Q: What happens if k is equal to total number of training points?

→All points are classified as the largest class, point's coordinates in feature space do not matter anymore

K-nearest neighbor classifier

- In theory error will first decrease with k, then increase again
- May not see such a curve on real data

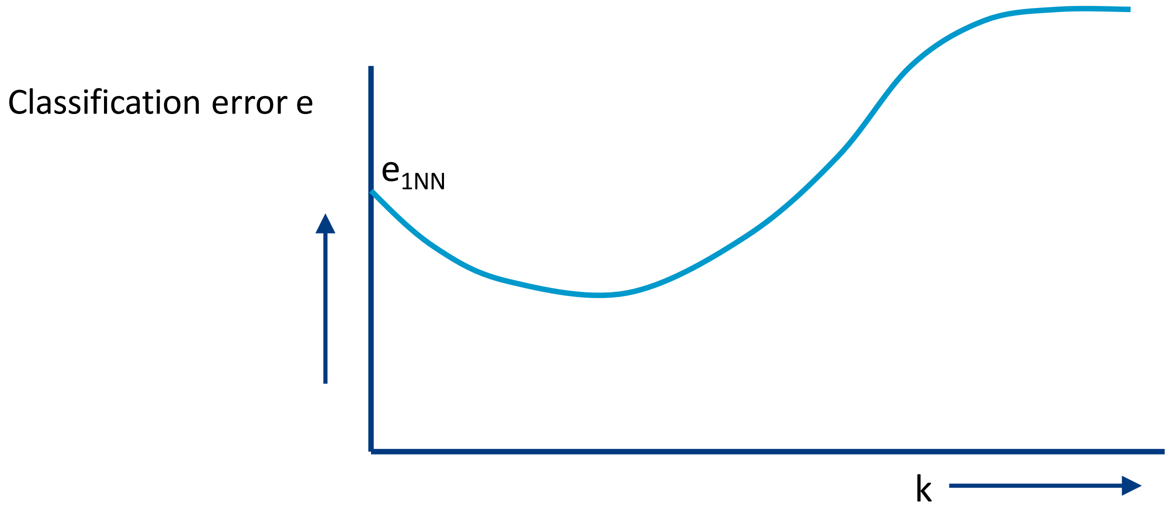


Image source: David Tax

K-Means vs K-NN

What is similar / different between k-Means and k-NN?

- Clustering vs classifier
- Linear vs non-linear
- Optimization vs lazy
- Both have K parameter
- Both based on distances

Histogram-based classifier

Idea:

- Split up the feature space into cells
- Decide class based on majority of points in each cell

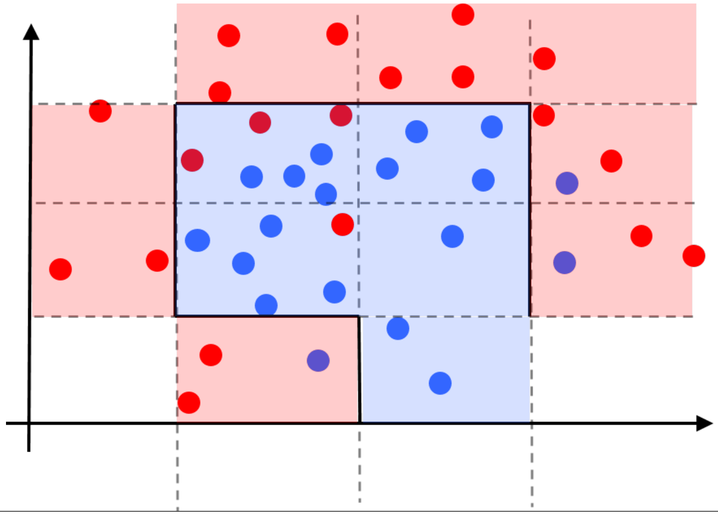
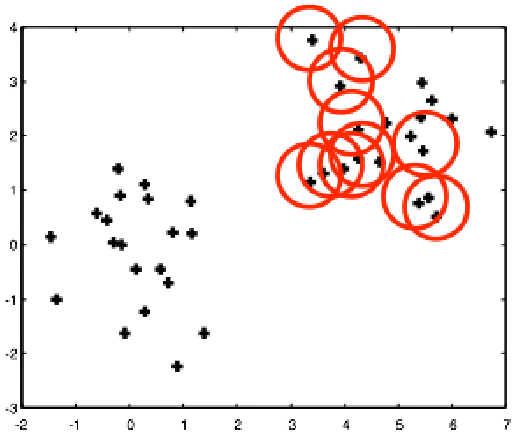


Image source: V. Cheplygina

Histogram-based → Parzen classifier

- Put Gaussian cells on locations of data points
- Sum of Gaussians = density for the data



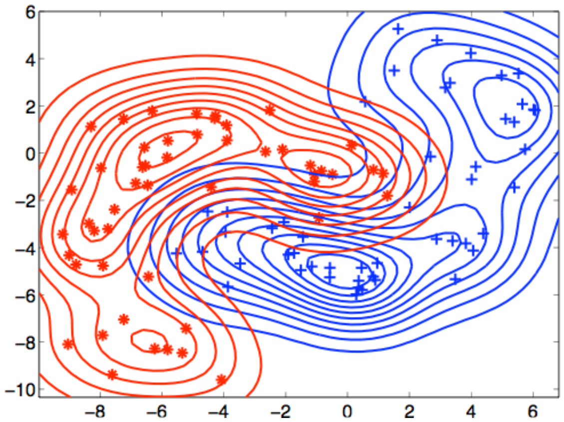
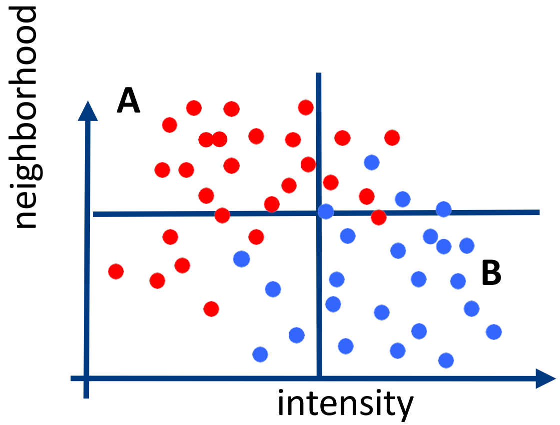


Image source: V. Cheplygina

Decision tree classifier

Find rules that split up the data as well as possible



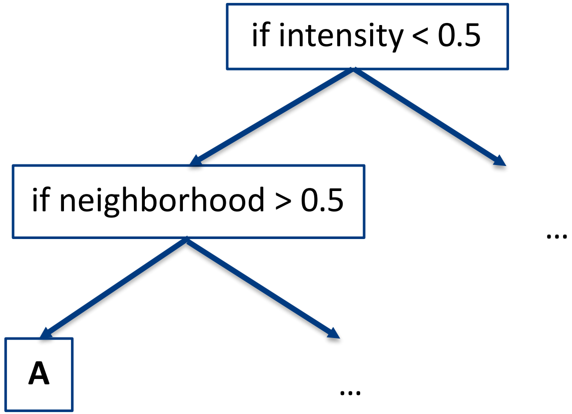
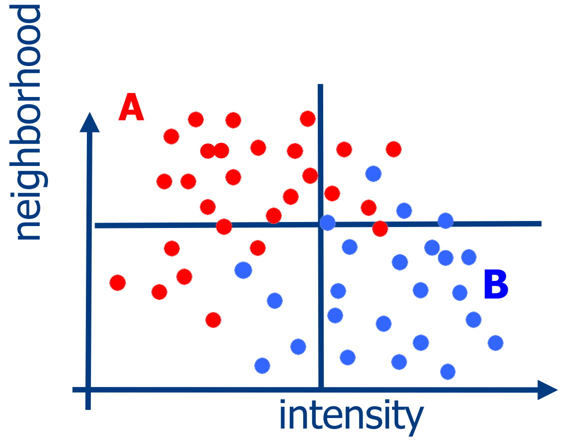


Image source: V. Cheplygina

Decision tree classifier

Until some stopping criterion:

- Find the feature that splits up the data “the best”
- Add another split etc



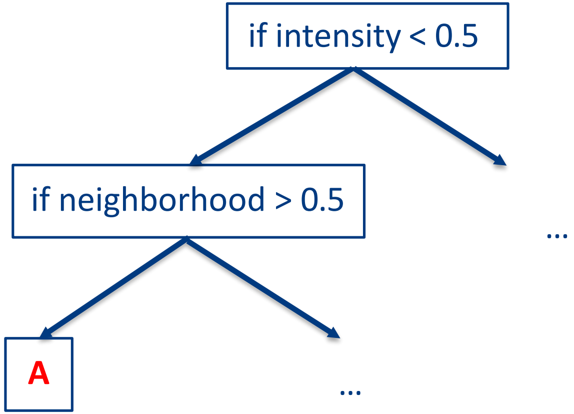
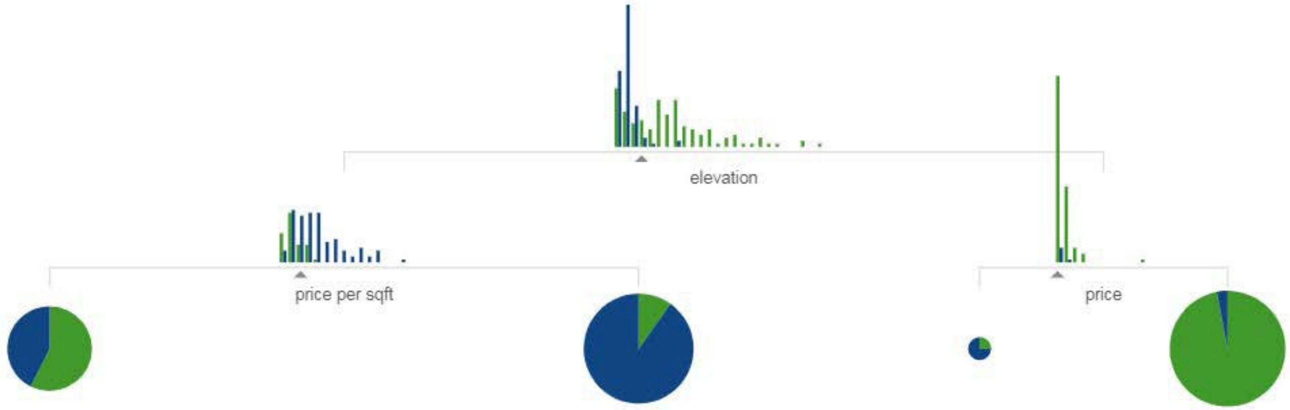


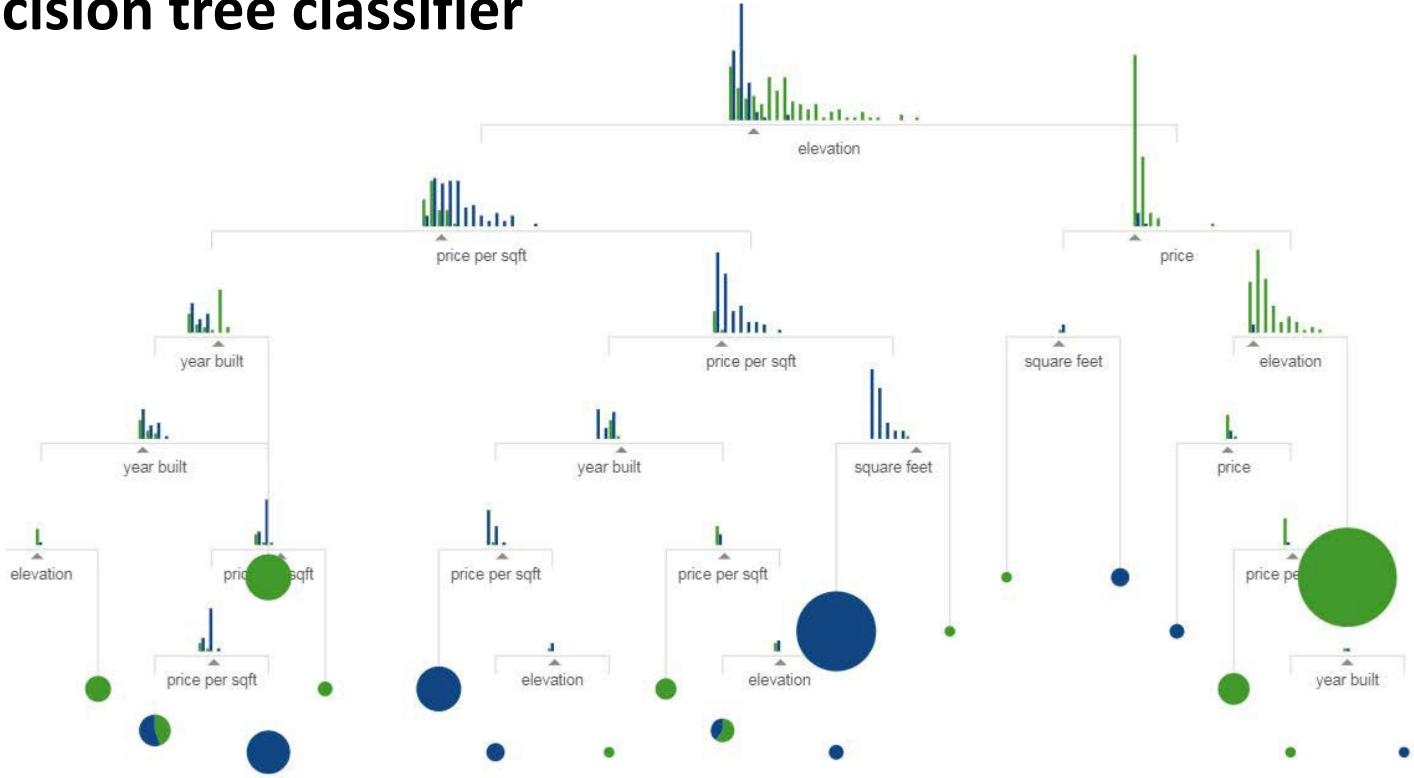
Image source: V. Cheplygina

Decision tree classifier



Further read: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Decision tree classifier

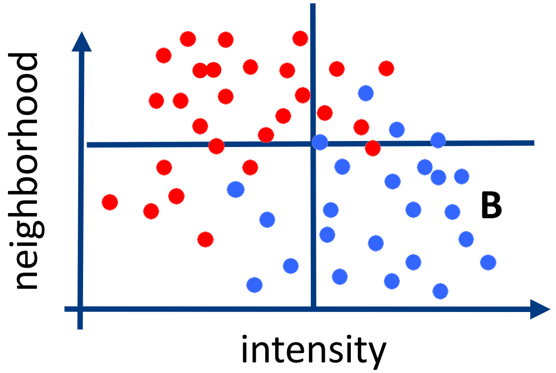


http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Decision tree classifier

Different stopping criteria

- Continue until each end point (leaf node) is from the same class
- Continue until maximum depth (not all data classified correctly!)



Outline

- Recap of previous lecture
- Segmentation by clustering
- Segmentation by classifiers
- Generalization and overfitting

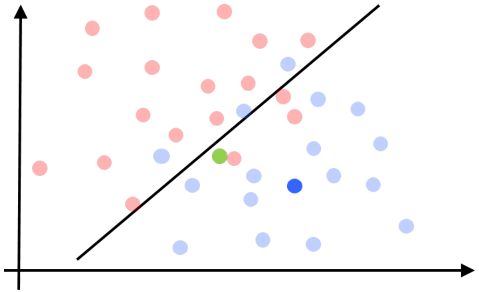
Generalization

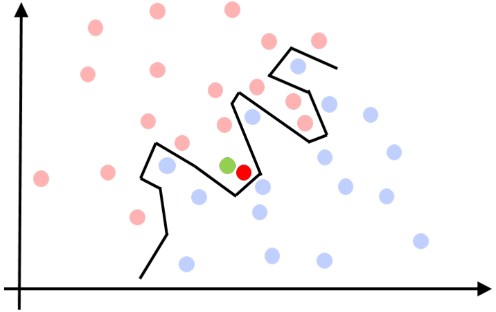
Which classifier do you prefer?

Metric: Correct for … % voxels?

Nearest mean: 90%

Nearest neighbor: 100%





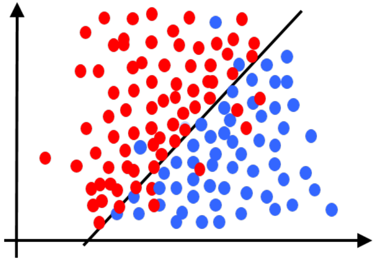
Generalization

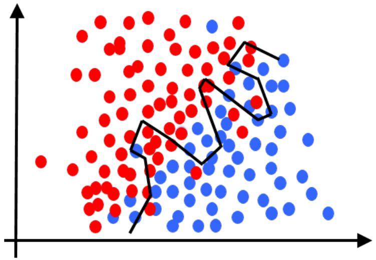
Which classifier do you prefer?

Metric: Correct for … % of future pixels?

Nearest mean: 90%

Nearest neighbor: 75%



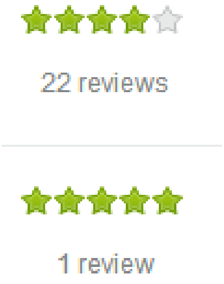


Generalization

- Current pixels = training set
- Future pixels = test set
- Training error can be low, but this doesn't mean the test error will be low as well!
- **Which factors influence this?**

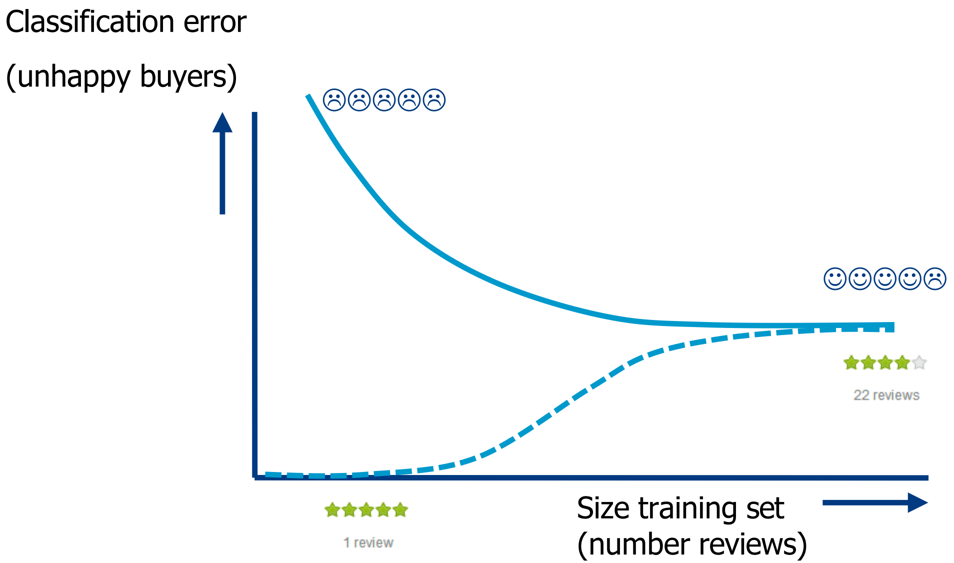
Generalization example

- Previous buyers: training set
- Future buyers (you): test set
- Which phone do you buy?



Generalization example

Based on 1 review with 5 stars, possible that future buyer will be unhappy (i.e. the buyer made an error)



IMAG/e TU/e

Generalization

- The classifier has learned the **true patterns** in the training data
- These patterns also **apply** to the test data
- The classifier can create **good predictions** for the test data

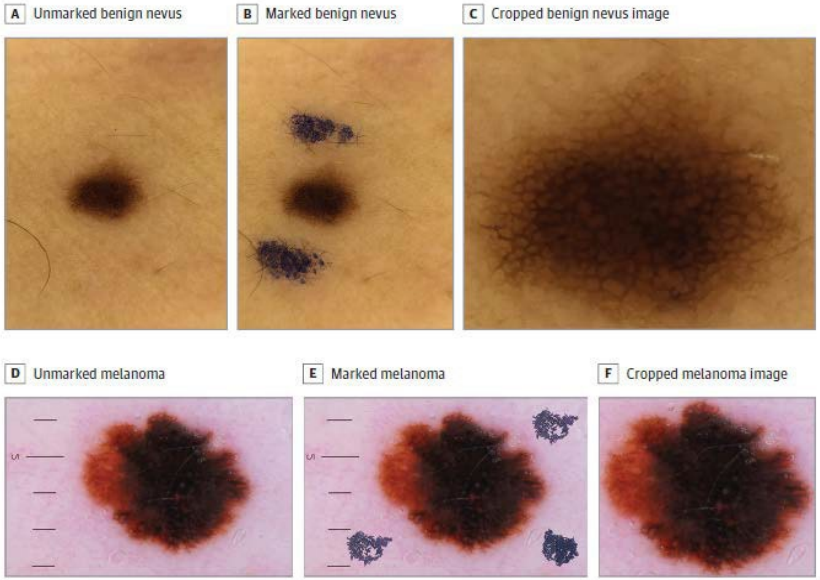
Overfitting

- Opposite of generalization
- The classifier has learnt patterns which separate the training data (training error is low)
- These are not the true patterns / they do not separate the test data
- The classifier can **not** create good predictions for the test data (test error is high)

14

Overfitting example

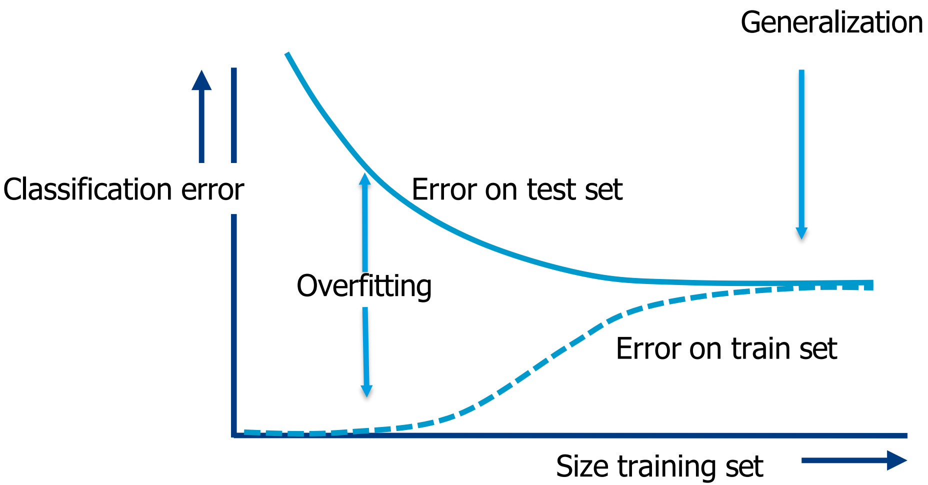
Figure 1. Convolutional Neural Network (CNN) Classification and Melanoma Probability Scores for Dermoscopic Images of Unmarked, Marked, and Cropped Benign Nevus and Melanoma



A gentian violet surgical skin marker was used to highlight the marked examples. A, CNN classification: benign; score, 0.001. B, CNN classification: malignant; score, 0.981. C, CNN classification: benign; score, 0.001. D, CNN classification: malignant; score, 0.999. E, CNN classification: malignant; score, 0.999. F, CNN classification: malignant; score, 0.999.

https://jamanetwork.com/journals/jamadermatology/article-abstract/2740808

Generalization vs Overfitting



Causes of overfitting

Overfitting when

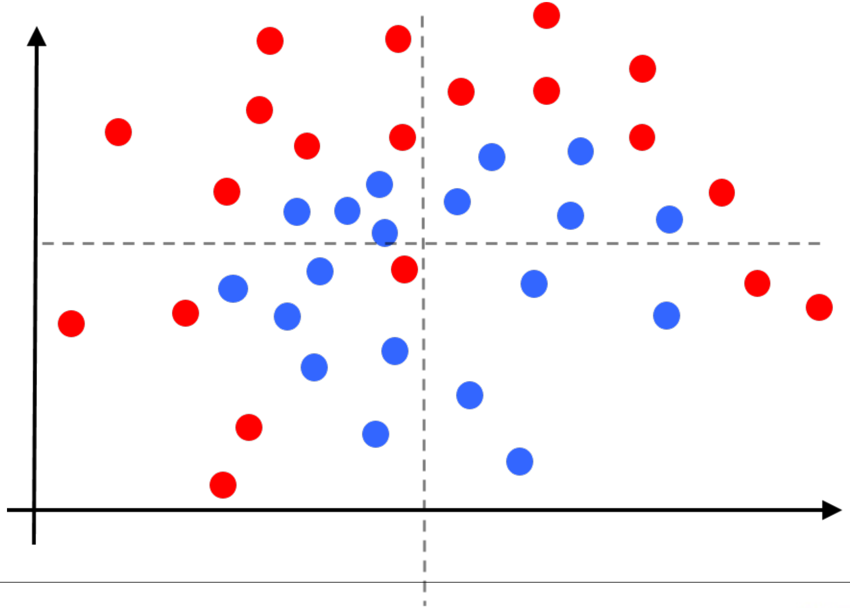
1) Too few representative training examples

2) …

Overfitting example

We use the “histogram-based classifier”

Divide space into cells, label each cell by majority of samples

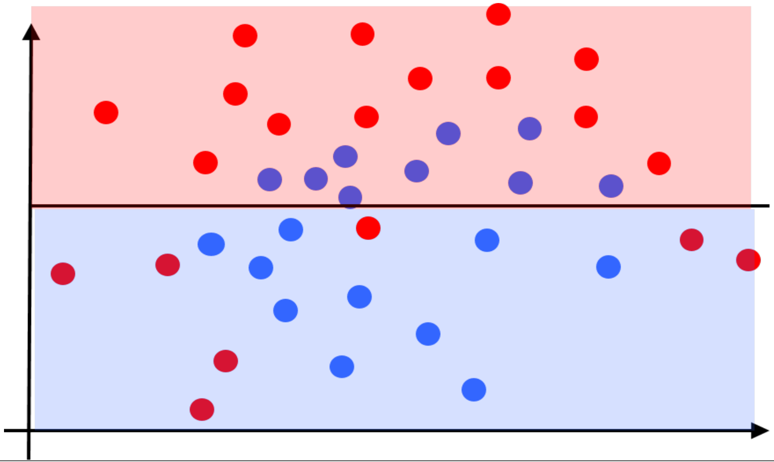


Overfitting example

Number of bins determines how simple/complex the model is

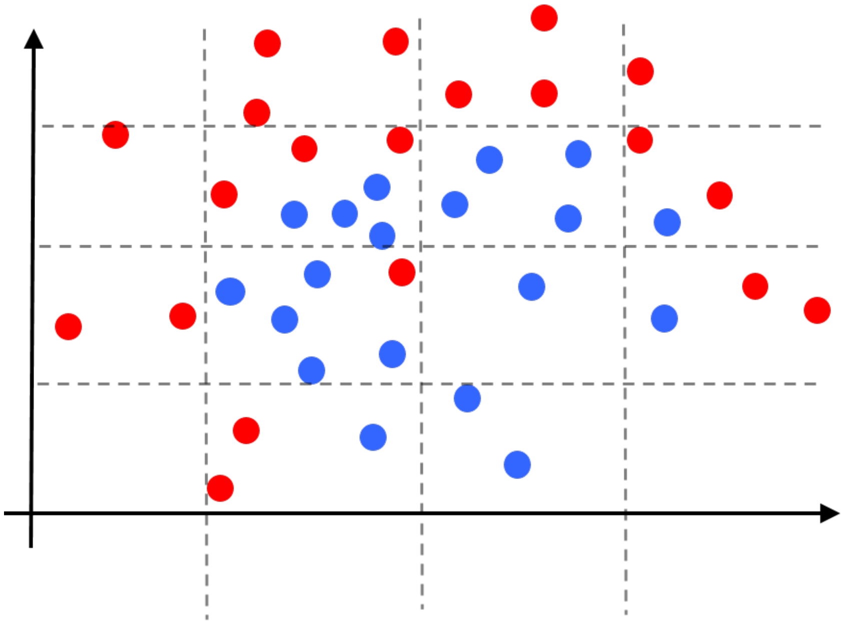
Underfitting:

- model is too simple, cannot capture the patterns in the data
- High training error, high test error



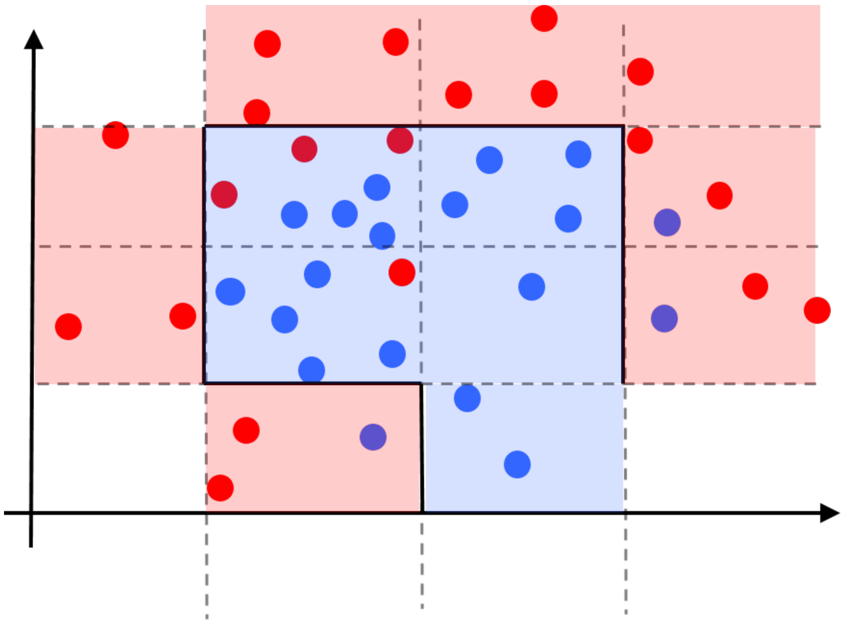
Overfitting example

We can increase complexity (and fit the data better) by adding more cells



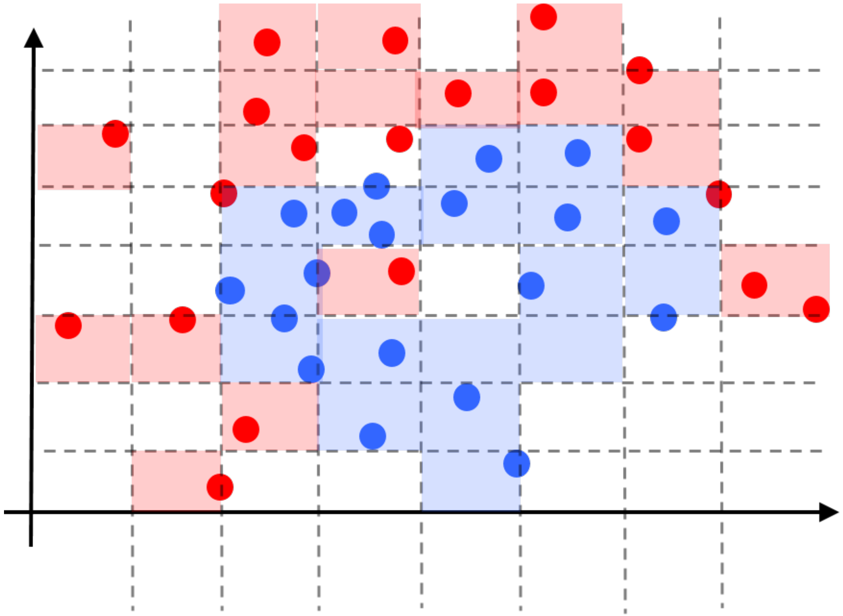
Overfitting example

This is a better model, but there are still some errors



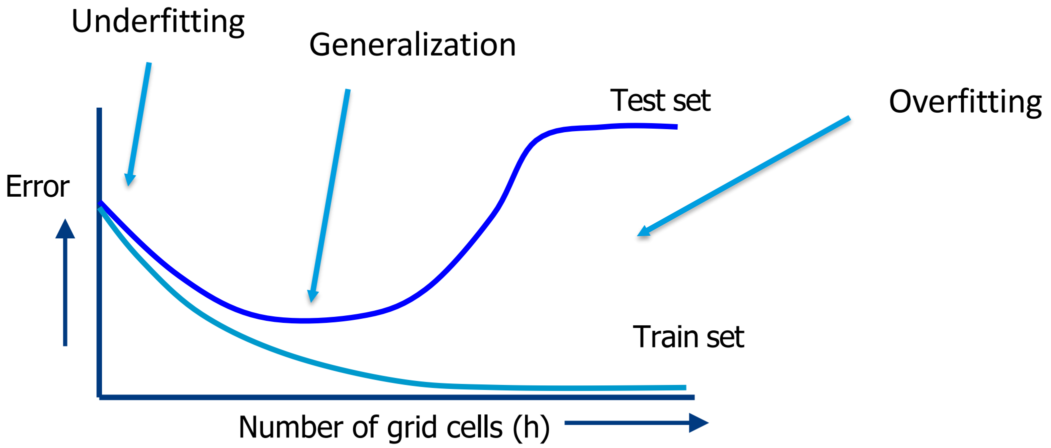
Overfitting example

Further increasing complexity will fit training data perfectly, but the decisions will be noisy (e.g. red cell in the middle of blue), many empty cells – **overfitting** has occured



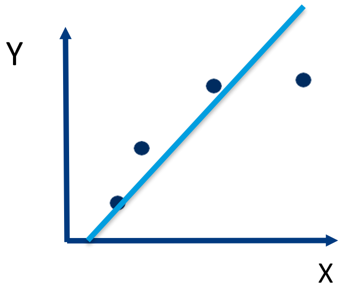
Overfitting

More cells = more complexity = can more easily capture patterns in data

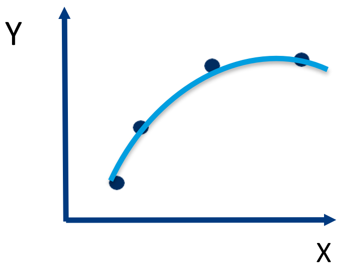


Generalization/overfitting example

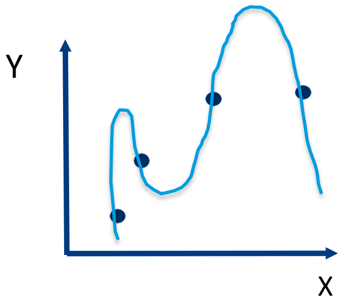
- NOTE: Now there is only 1 feature (X), the y-axis shows the output variable
- When training a classifier, we are searching for a function f, such that f(X) = Y
- Function in this case is a polynomial



$Y = aX + b$



$Y = aX^2 + bX + c$



$Y = aX^4 + bX^3 + cX^2 + dx + e$

More explanations: https://www.youtube.com/watch?v=UJUYCywWsek

Causes of overfitting

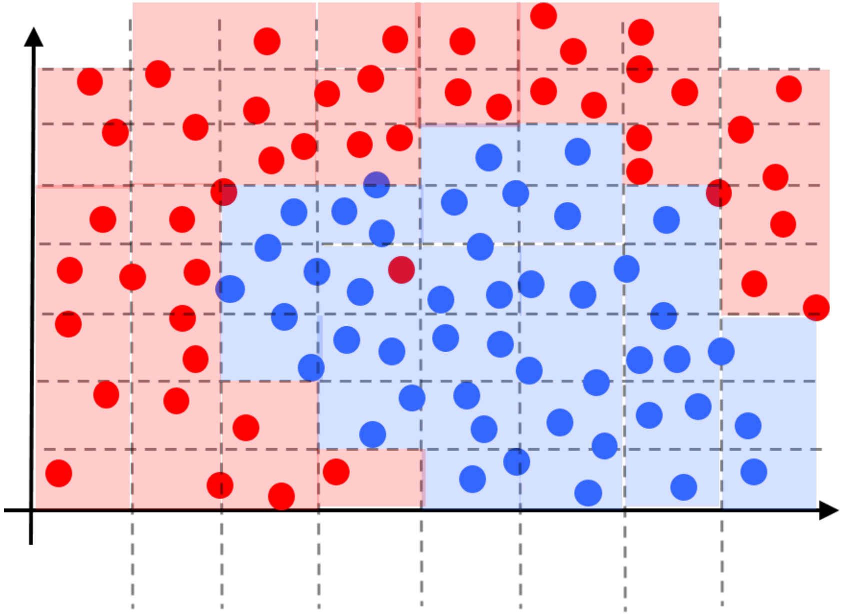
Overfitting when

1) Too few representative training examples

2) Too complex classifier +

Generalization/overfitting

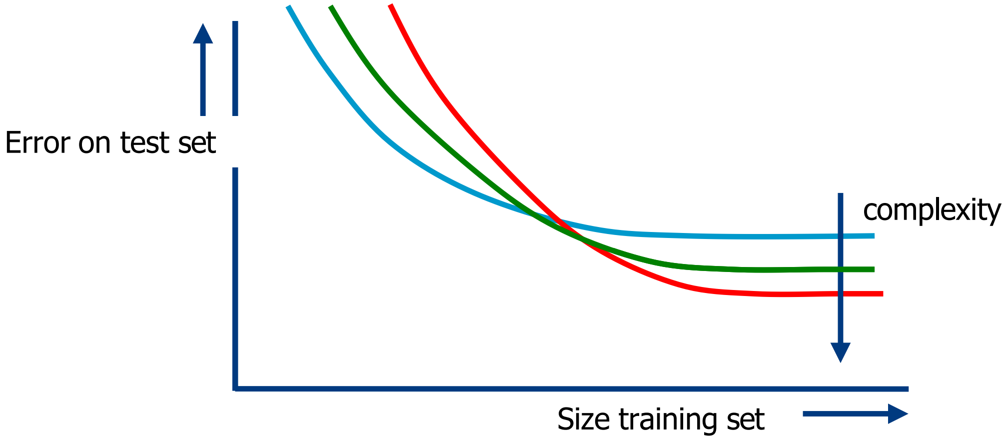
Note: The amount of complexity the classifier can, depends on the amount of data



Generalization/overfitting

Trade-off sample size and classifier complexity

- Use **simple** classifiers (e.g. nearest mean or another linear classifier) when data is **small**



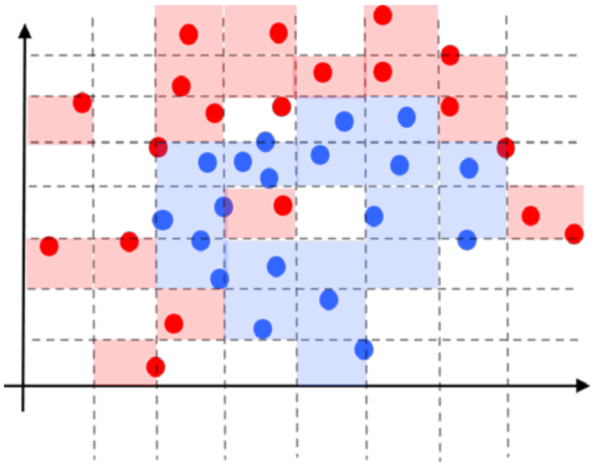
Nearest mean could be the blue classifier, 1-nearest neighbor the red one

Causes of overfitting

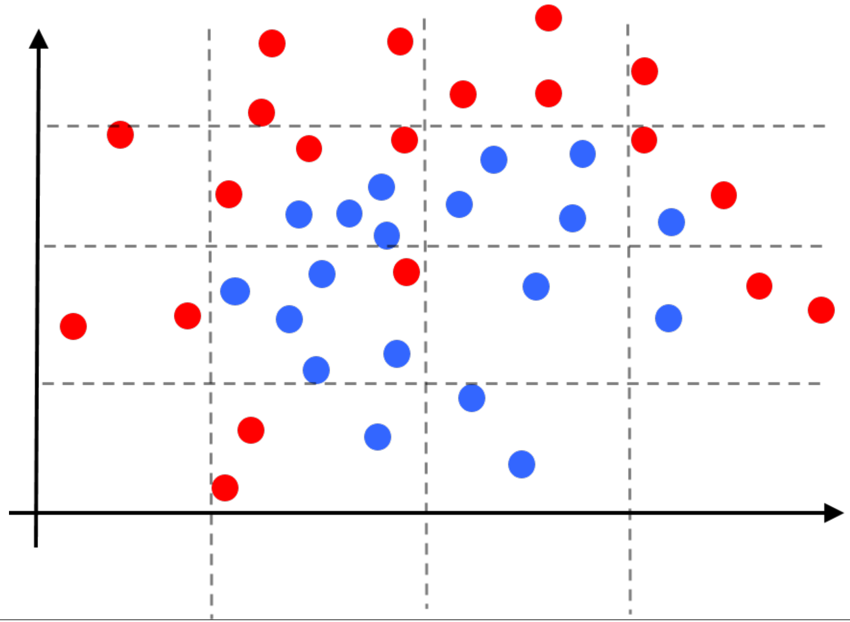
What if instead of adding more compexity to the model, we add more features, such that data is linearly separable?

Imagine we add a third feature, “how close is this point to the center of the data”

Blue points will have lower values on this feature (closer to center) – can linearly separate the data again



- BUT, in 2D we need at least 16 samples to fill each cell
- If we add a third dimension, we need more samples



Causes of overfitting

Overfitting when

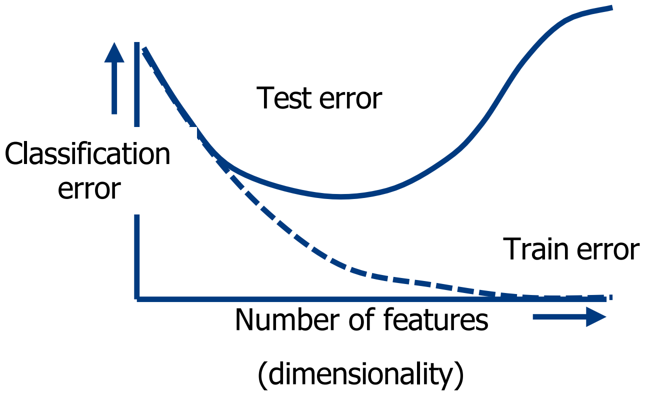
1) Too few representative training examples

2) Too complex classifier +

3) Too many features ("Curse of dimensionality")

"Curse of dimensionality"

The more dimensions you add, the "emptier" the space becomes, the easier it is to perfectly fit a model to the training data



Further reading: http://www.visiondummy.com/2014/04/curse-dimensionality-affect- classification/

“Curse of dimensionality”

Consider 100 uniformly distributed points in a unit line / square / cube …

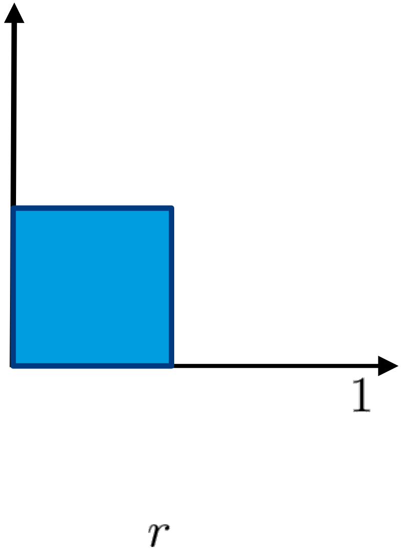
Your goal is to find 1 nearest neighbor (1% of the points, V = 0.01)

How far (what is the value of r) do you need to travel on a line? (m = 1)

How far to travel on a square? (m = 2)

…

$$r^m = V$$



Dimensionality reduction

Feature selection

- Recursive feature elimination
- Forward/backward feature addition

Feature extraction

- Principal component analysis
- Manifold learning (t-SNE)

Important note

We can “diagnose” generalization or overfitting by looking at the training error and test error.

But by looking at the test error, the test data becomes part of the training data!

Do not touch test data until the very end, split training data into “training-training” and “training-test” (also called validation)

Summary

The student can:

- Apply **K-Means** by hand on a small dataset.
- Apply the nearest mean, nearest neighbor classifier or similar simple **classifiers** by hand on a small dataset
- Describe **properties** of simple classifiers (type of boundary, sensitivity to scaling, sensitivity to irrelevant features)
- Explain the **effect of classifier** properties on the result of a segmentation.

- Explain what is meant by **generalization and overfitting**
- **Diagnose** generalization/overfitting given scatterplot of dataset, decision boudnary and/or classifier results (error plots)
- **Compare** classifiers based on their complexity / ability to overfit, understand how choice of parameters influences complexity
- Name **reasons overfitting** can occur / be able to suggest strategies to reduce overfitting

Further reading

- Book: Guide to medical image analysis: https://link.springer.com/book/10.1007/978-1-4471-2751-2

- Online material: http://www.r2d3.us/visual-intro-to-machine-learning-part-2/ - overfitting

Thank you

r.su@tue.nl

Next: Graph cuts for image segmentation