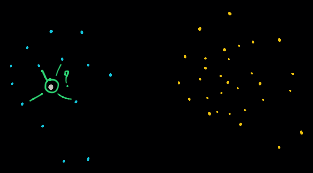
kNN for binary classification



→ Bunch of points.

→ Look at surrounding points to predict the class of a new point.

→ Find k nearest neighbours

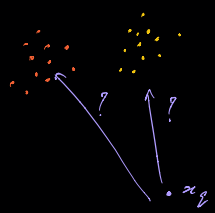
→ majority or average or median, etc.

→ Phone and Skype interviews won't have whiteboard. So it's important to explain stuff without pen/paper in a simple concise way

Pretty basic most of the time.

Where does kNN mess up?

If x_q (query pt.) is too far. kNN won't be reasonable.



kNN doesn't work with high-dim data. because Euclidean distance fails at the level. Nothing wrong with kNN.

All depends on distance metric.

Very high space & time complexity.

If dataset has outliers, it fails.

→ There's a threshold of stuff you should remember.

→ It's okay to not remember syntax of obscure fn.

→ It's NOT fucking okay to forget syntax of for loop or dictionary.

A lot of the questions are already discussed in the course

Euclidean → $(x_2 - x_1)^2$ accentuates the differences.

So when it's high-dim, still be a problem. especially with outliers.