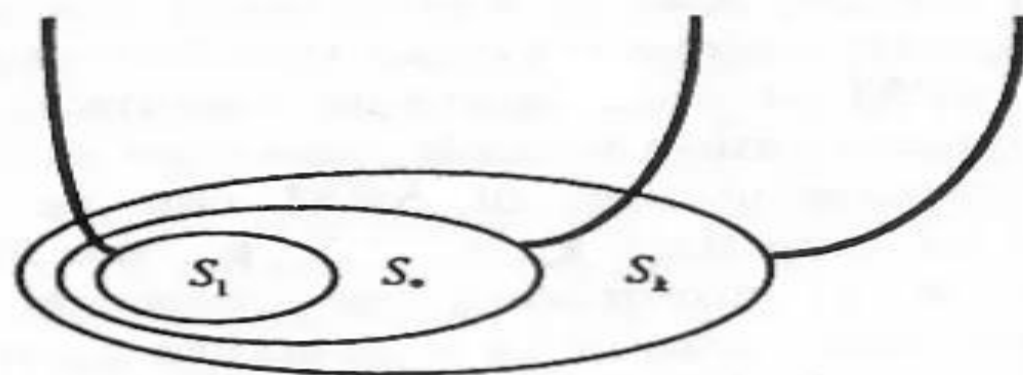
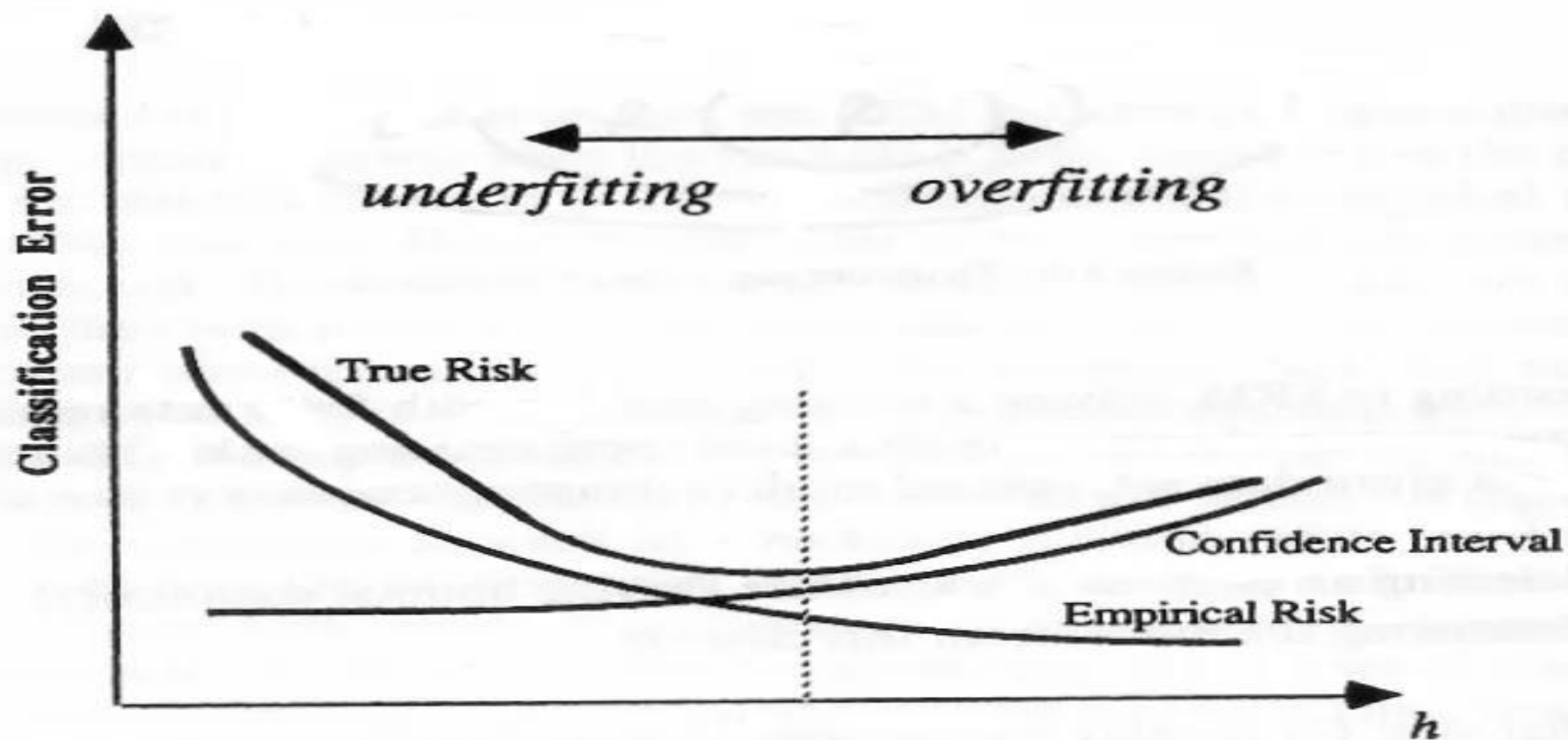


In Sample Vs. Out of Sample Error

- **In Sample Error**: The **error** rate that we get on the same data set (i.e. training) we used to build our predictor and it is also called as re-substitution **error**.
- **Out of Sample Error**: The **error** rate that we get on a new data set (i.e. testing) and it is also called as generalization **error or expected loss or risk**. Basically, it measures the accuracy of an algorithm to predict outcome values for previously unseen data.

Vapnik-Chervonenkis (VC) Dimension

- The VC dimension is defined by Vapnik and Chervonenkis. It is the cardinality (size) of the largest set of points that the classification algorithm can shatter.
- The VC dimension of the hypothesis space H , $VC(H)$, is the size of the largest finite subset of the instance space X that can be shattered by H .
- The **high VC dimension allows for a lower empirical risk** (sample data error), **but also introduces a higher confidence interval**. This interval can be seen as the confidence in the model's ability to generalize.
- **It is a measure of the capacity (complexity, expressive power, richness, or flexibility) of a set of functions that can be learned by a statistical binary classification algorithm.**



How do we characterize “power”?

- Different machine learning models have different amounts of “power”.
- Tradeoff between:
 - More power: Can model more complex classifiers but might overfit.
 - Less power: Not going to overfit, but restricted in what it can model.
- How do we characterize the amount of power?

Vapnik-Chervonenkis Dimension

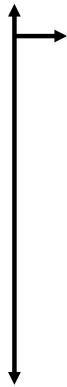
Expressive power, or **capacity**, of a **hypothesis class**

- Linear classifiers in d -dimensional space
 - Degree k polynomial classifiers
 - Hierarchical axis-parallel classifiers (decision trees)
- Measured by ability of hypothesis class to **shatter** n points

Shattering

Classify points into all possible labels

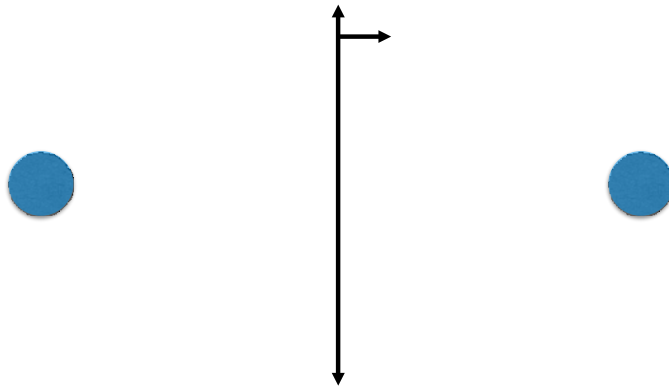
✓
++, +-, -+, --



Shattering

Classify points into all possible labels

\checkmark
++ , +- , -+ , --



Shattering

Classify points into all possible labels

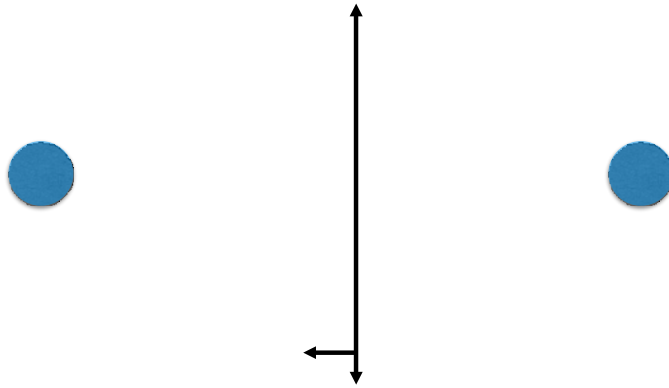
\checkmark \checkmark \checkmark
++, +-, -+, --



Shattering

Classify points into all possible labels

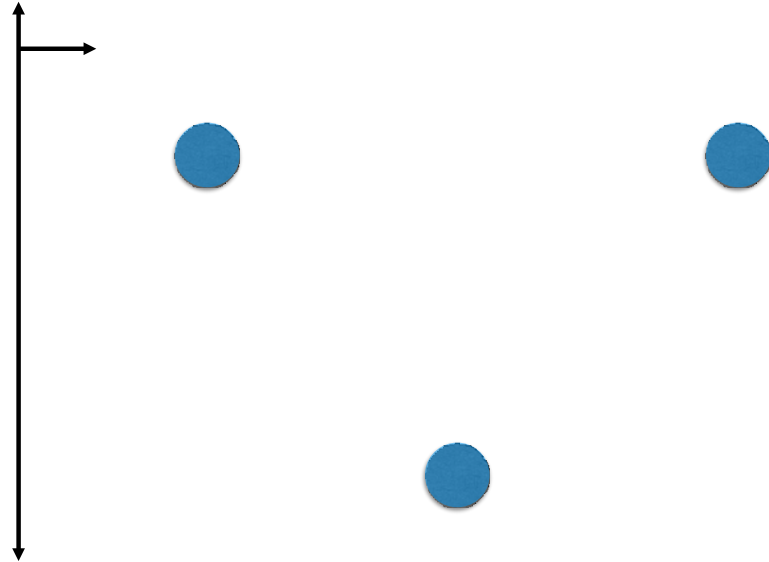
\checkmark \checkmark \checkmark \checkmark
++, +-, -+, --



Shattering

Classify points into all possible labels

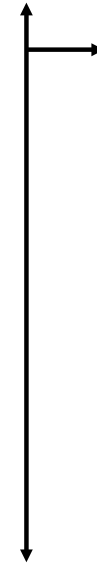
✓
+++, ++-, +-+, +- -, -++-, - + -, - - -, - - +



Shattering

Classify points into all possible labels

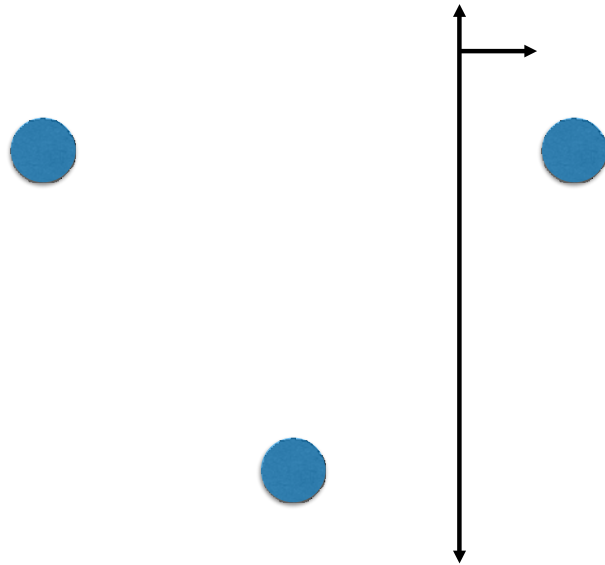
[✓]
+++, ++-, +-+, +- -, -++-, - + -, [✓] - - -, - - +



Shattering

Classify points into all possible labels

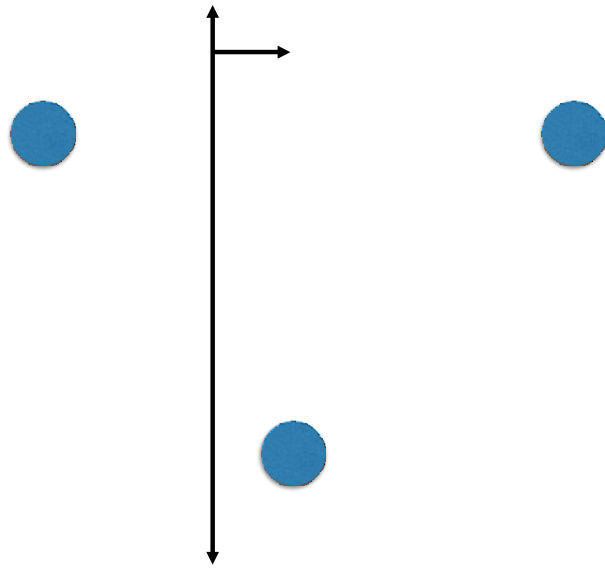
\checkmark \checkmark \checkmark
+++, ++-, +-+, +- -, -++-, - + -, - - -, - - +



Shattering

Classify points into all possible labels

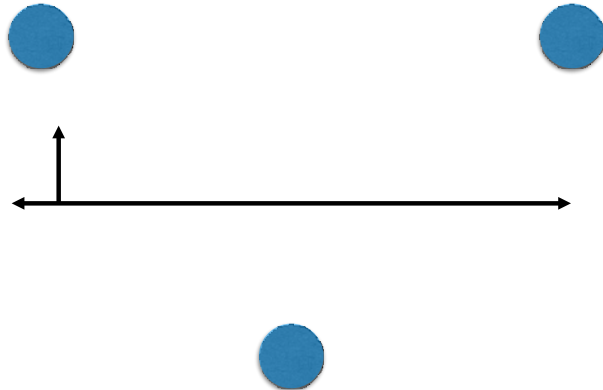
\checkmark \checkmark \checkmark \checkmark
+++, ++-, +-+, +- -, -++-, -+-, ---, ---+



Shattering

Classify points into all possible labels

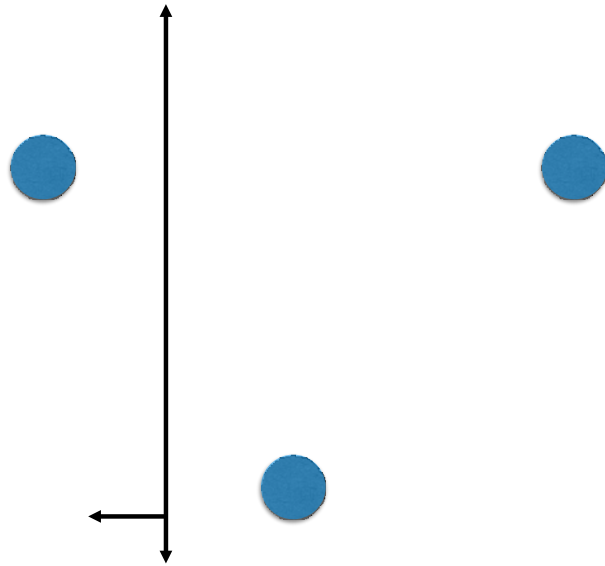
\checkmark \checkmark \checkmark \checkmark \checkmark \checkmark
+++, ++-, +-+, +- -, -++-, -+-, ---, ---+



Shattering

Classify points into all possible labels

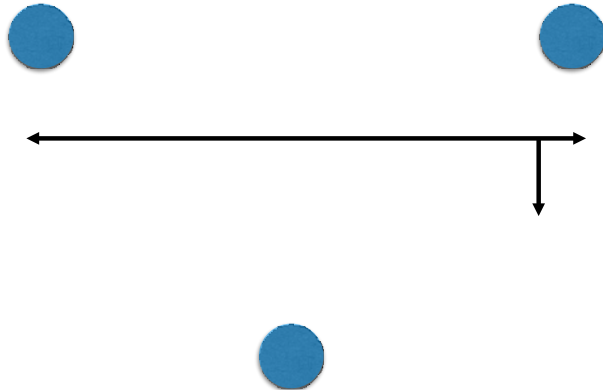
✓ ✓ ✓ ✓ ✓ ✓
+++ , ++- , +-+ , +- - , -++ , - + - , - - - , - - +



Shattering

Classify points into all possible labels

✓ ✓ ✓ ✓ ✓ ✓ ✓
+++, ++-, +-+, +- -, -++-, - + -, - - -, - - +



Shattering

Classify points into all possible labels



4 points cannot be shattered by 2d linear classifier

VC Dimension

- VC dimension of hypothesis class H :
- Maximum number of examples that can be shattered by H
- **Examples can be arranged (feature values) in any way**
- **Must be shattered in same arrangement**
- In general: linear classifier has VC dimension $(d + 1)$