Machine Learning - Lecture 11: VC dimension

Chris Thornton

November 9, 2011

Introduction

Decision-tree learning is typical of machine learning in its approach to model development.

Each time a split is introduced, the model is made a bit more detailed.

More training cases covered.

The model is said to be **incrementally refined** using the data as a source of reference.

The problem of over-fitting

Unfortunately, this approach produces problems if there are errors in the data (which there usually are).

In this case there is the danger of reaching a point where refinements are simply 'learning the errors'.

This is known as **over-training** or **over-fitting**.

The U-shaped peformance curve

Fortunately, it is possible to detect over-fitting.

As model refinement continues we expect to see an improvement in performance on both seen and unseen data.

If the data contains some errors, there will come a point where refinements are simply modeling errors.

At this point, we should see deteriorating generalization, even while performance on seens continues to improve.

Cross-validation to the rescue

It is with this problem that cross-validation comes into its own.

We can use cross-validation methods to *detect* the point at which refinements appear to be producing the effect of over-fitting.

The idea is to terminate learning as soon as we go past the point where performance on unseen examples starts to deteriorate.

Unfortunately, in some situations, we see quite significant variations in generalization performance *prior* to any over-fitting.

Spotting the critical moment can be quite challenging.

Bias-based strategies

Another approach to the problem of over-fitting involves control of bias.

More strongly biased methods are more limited in terms of the patterns they can represent.

So another way of making sure we don't end up 'training on noise' is to use a method whose bias effectively rules such patterns out.

In practice, this may be hard to achieve if we don't know what the errors are.

However, the general rule applies.

Bias-based strategies

Another approach to the problem of over-fitting involves control of bias.

More strongly biased methods are more limited in terms of the patterns they can represent.

So another way of making sure we don't end up 'training on noise' is to use a method whose bias effectively rules such patterns out.

In practice, this may be hard to achieve if we don't know what the errors are.

However, the general rule applies.

- A stronger bias implies less vulnerability to over-fitting.

Bias-based strategies

Another approach to the problem of over-fitting involves control of bias.

More strongly biased methods are more limited in terms of the patterns they can represent.

So another way of making sure we don't end up 'training on noise' is to use a method whose bias effectively rules such patterns out.

In practice, this may be hard to achieve if we don't know what the errors are.

However, the general rule applies.

- A stronger bias implies less vulnerability to over-fitting.
- A weaker bias implies more vulnerability.

Bias-based strategies

Another approach to the problem of over-fitting involves control of bias.

More strongly biased methods are more limited in terms of the patterns they can represent.

So another way of making sure we don't end up 'training on noise' is to use a method whose bias effectively rules such patterns out.

In practice, this may be hard to achieve if we don't know what the errors are.

However, the general rule applies.

- A stronger bias implies less vulnerability to over-fitting.
- A weaker bias implies more vulnerability.

We want to achieve the strongest bias possible, but in a way that still allows significant patterns to be identified and represented.

VC definition

VC dimension is a formal measure of bias which has played an important role in mathematical work on learnability.

The VC dimension of a representation system is defined to be

VC definition

VC dimension is a formal measure of bias which has played an important role in mathematical work on learnability.

The VC dimension of a representation system is defined to be

- the maximum number of datapoints that can be separated (i.e., grouped) in *all* possible ways.

VC definition

VC dimension is a formal measure of bias which has played an important role in mathematical work on learnability.

The VC dimension of a representation system is defined to be

- the maximum number of datapoints that can be separated (i.e., grouped) in *all* possible ways.

Another way of saying this is to describe it as the the most datapoints that can be 'shattered' by the representation.

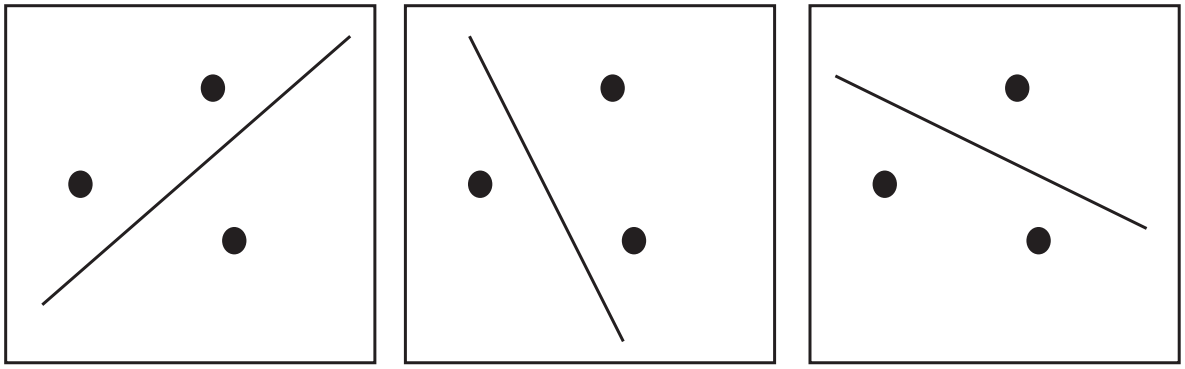
More powerful representations are able to shatter larger sets of datapoints. These have higher VC dimension.

Less powerful representations can only shatter smaller sets of datapoints.

These then have lower VC dimension.

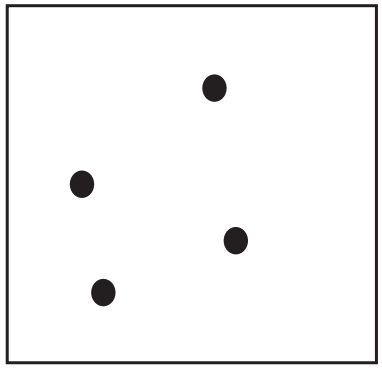
Classic VC-dimension illustration

3 datapoints can be grouped in all possible ways



Model based on single linear division

4 datapoints cannot



VC dimension = 3

VC dimension as a definition of bias

VC dimension seems to focus on a particularly demanding representation task, i.e., representing all possible ways of grouping datapoints.

From the intuitive point of view, this makes it less than ideal as a general measure of bias strength.

We could have a system with very low VC dimension that is actually quite weakly biased. This would happen, for example, if the system was able to *almost* shatter large datasets, while only being able to *fully* shatter very small ones.

VC dimension in mathematics

VC dimension is useful in formal analysis of learnability, however.

This is because VC dimension provides an *upper bound* on generalization error.

The mathematics of this are quite complex.

The basic idea is that reducing VC dimension has the effect of eliminating potential generalization errors.

So if we have some notion of how many generalization errors are possible, VC dimension gives an indication of how many could be made in any given context.

The subfield of **Computational Learning Theory** is concerned with deriving VC-dimension bounds in different training scenarios.

Summary

Summary

- With incremental model refinement we have the risk of ‘training on noise’, i.e,. over-fitting.

Summary

- With incremental model refinement we have the risk of 'training on noise', i.e,. over-fitting.
- A system that is over-fitting will tend to show deteriorating generalization.

Summary

- With incremental model refinement we have the risk of 'training on noise', i.e,. over-fitting.
- A system that is over-fitting will tend to show deteriorating generalization.
- Can address the problem through cross-validation strategies which aim to identify the point at which generalization starts to deteriorate.

Summary

- With incremental model refinement we have the risk of 'training on noise', i.e,. over-fitting.
- A system that is over-fitting will tend to show deteriorating generalization.
- Can address the problem through cross-validation strategies which aim to identify the point at which generalization starts to deteriorate.
- Can also address the problem through strenghthening bias.

Summary

- With incremental model refinement we have the risk of 'training on noise', i.e,. over-fitting.
- A system that is over-fitting will tend to show deteriorating generalization.
- Can address the problem through cross-validation strategies which aim to identify the point at which generalization starts to deteriorate.
- Can also address the problem through strenghthening bias.
- VC dimension is a formal measure of bias that has been very useful in formal learning theory.

Summary

- With incremental model refinement we have the risk of 'training on noise', i.e,. over-fitting.
- A system that is over-fitting will tend to show deteriorating generalization.
- Can address the problem through cross-validation strategies which aim to identify the point at which generalization starts to deteriorate.
- Can also address the problem through strenghthening bias.
- VC dimension is a formal measure of bias that has been very useful in formal learning theory.

Questions

Questions

- How does the problem of over-fitting arise in the case of k-means clustering?

Questions

- How does the problem of over-fitting arise in the case of k-means clustering?
- How does the problem of over-fitting relate to the problem of lookup tables?

Questions

- How does the problem of over-fitting arise in the case of k-means clustering?
- How does the problem of over-fitting relate to the problem of lookup tables?
- Consider a representational framework based on centroids in a 2d space. What is the VC dimension?

Questions

- How does the problem of over-fitting arise in the case of k-means clustering?
- How does the problem of over-fitting relate to the problem of lookup tables?
- Consider a representational framework based on centroids in a 2d space. What is the VC dimension?