Machine Learning

Stephen G. Ware

CSCI 4525 / 5525

Narrative Intelligence LAB

THE UNIVERSITY of NEW ORLEANS

Learning

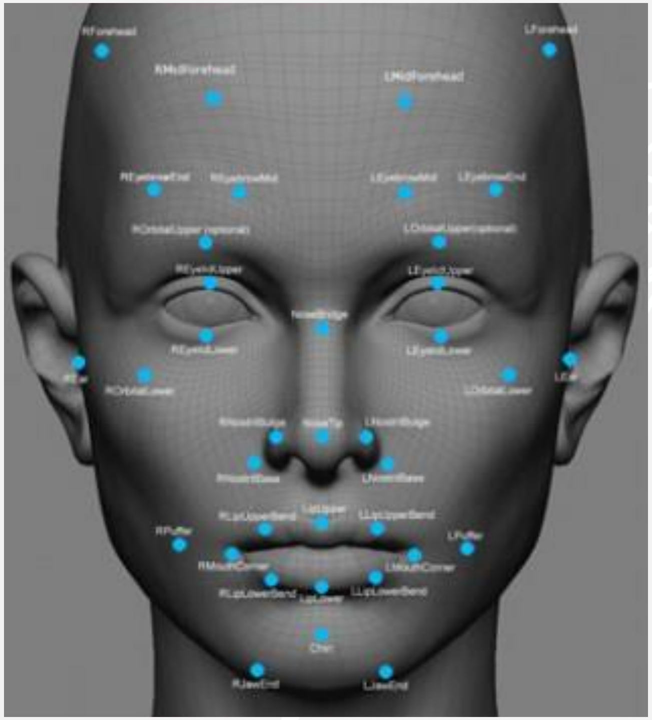
A agent which can **learn** is able to improve its performance based on observations. Learning is valuable for a number of reasons:

- Agent designers may not be able to anticipate all possible situations
- Agent designers may not be able to anticipate changes in the task over time.
- Agent designers may not know how to solve a given problem

Facial Recognition

Consider the task of facial recognition. While human programmers may be good at recognizing faces, they had a very difficult time designing programs for that task.

However, they were able to design agents which could learn to recognize faces.



Reasoning in AI

There are at least three types of reasoning used by difference kinds of AI systems:

- Deductive Reasoning
- Abductive Reasoning
- Inductive Reasoning

All can be framed in terms of causal reasoning:

$$\underbrace{a}_{\text{antecedent}} \rightarrow \underbrace{b}_{\text{consequent}}$$

Deductive Reasoning

Given a general rule and some specific antecedent, **deductive reasoning** applies the rule to conclude the consequent.

Given: $a \rightarrow b$ and a, conclude b.

Example: All birds can fly. Polly the parrot is a bird. Therefore, Polly can fly.

Most systems based on formal logic, such a theorem provers and planners, use deductive reasoning.

Abductive Reasoning

Given a general rule and some specific consequent, **abductive reasoning** uses the rule to conclude the antecedent.

Given: $a \rightarrow b$ and b, conclude a.

Example: All birds can fly. Polly the parrot can fly. Therefore, Polly is (probably) a bird.

Most probabilistic inference systems, such as Bayes Nets, use abductive reasoning.

Inductive Reasoning

Given a specific antecedent and consequent, **inductive reasoning** uses the evidence to conclude a rule.

Given: a and b, conclude $a \rightarrow b$.

Example: All the birds we have seen can fly. Therefore, all birds are (probably) able to fly.

Most machine learning systems use inductive reasoning.

Machine Learning

Machine learning is the process of using examples of input (data) to construct a hypothesis (model) which can predict the correct output.

Machine learning tasks are usually divided into three categories:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Given a set of training data, a **supervised learning** approach constructs a model from the training data in the hopes of using that model to make accurate predictions about new data.

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Parrot	Yes	Yes
3	Slamon	No	No
4	Hawk	Yes	Yes
5	Flying Squirrel	Yes	No
6	Ostrich	No	Yes
7	Sparrow	Yes	Yes

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Eagle	Yes	?
3	Squid	No	?
4	Penguin	No	?
5	Dog	No	?
6	Flying Fish	Yes	?
7	Finch	Yes	?

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Eagle	Yes	Yes
3	Squid	No	No
4	Penguin	No	No
5	Dog	No	No
6	Flying Fish	Yes	Yes
7	Finch	Yes	Yes

Unsupervised Learning

When no training data is available, **unsupervised learning** attempts to find useful patterns in data.

	A	B
1	**Common Name**	**Flies?**
2	Eagle	Yes
3	Squid	No
4	Penguin	No
5	Dog	No
6	Flying Fish	Yes
7	Finch	Yes

	A	B
1	**Group 1**	**Group 2**
2	Eagle	Squid
3	Flying Fish	Penguin
4	Finch	Dog
5		
6		
7		

Reinforcement Learning

Reinforcement learning refines its model over time based on periodic positive and negative reinforcement that tells it whether or not it is making correct predictions.

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Eagle	Yes	?
3	Squid	No	?
4	Penguin	No	?
5	Dog	No	?
6	Flying Fish	Yes	?
7	Finch	Yes	?

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Eagle	Yes	Yes
3	Squid	No	No
4	Penguin	No	No
5	Dog	No	No
6	Flying Fish	Yes	Yes
7	Finch	Yes	Yes

4 of 6 correct.

	A	B	C
1	**Common Name**	**Flies?**	**Bird?**
2	Eagle	Yes	Yes
3	Squid	No	No
4	Penguin	No	No
5	Dog	No	No
6	Flying Fish	Yes	No
7	Finch	Yes	Yes

5 of 6 correct.

Relationship to GOFAI

How is machine learning related to the good old fashioned AI (GOFAI)—i.e. representation and search?

Machine learning can be viewed as search through the space of models which explain a given data set.

However, most machine learning techniques do not employ the kind of search that we have discussed so far, so this relationship is tenuous.

Learned Models are Black Boxes

Even when learning is successful, it is often difficult to translate the learned model into intuitions which can be used by humans.

When this is true, we say that the model is a **black box**. We know that it works, but not why it works.

Most machine learning techniques make no claims about causality, only correlation. In other words, there may be a correlation between a and b, but we cannot claim that $a \rightarrow b$ or $b \rightarrow a$.

Supervised Learning Example

- For 14 days, you record information about the weather and whether or not it is a good day for playing Tennis.
- Your goal: Given information about the weather, determine if it is a good day to play Tennis.
- This is a supervised learning task because we start with labeled data.
- We will explore two simple classifiers: the majority classifier and Naïve Bayes.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
[illegible]	[illegible]	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Index

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Stro[illegible]	[illegible]
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Class Label

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	[illegible]	[illegible]	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attributes

Labeled Data

Should we include the index at an attribute when learning model?

No! The index is not a useful attribute of the data that we should expect to see repeated in future observations. In other words, Day 3 will never happen again, so including it in the model will not help.

This is a common rookie mistake in machine learning.

Evaluation

The most common way to evaluate the performance of a classifier is to divide the data into two parts: the **training set** and the **test set**.

The model is built using the training set, and its accuracy is evaluated using the test set.

Cross Validation

Traditionally, the data is divided into x number of equal sized groups called folds. The model is built and evaluated x times, each time with a different fold being used as the test set. The results of each run are then averaged together.

This approach is called ***x*-fold cross validation**.

When the fold size is 1, we call this special case **leave one out**.

Majority Classifier

One of the simplest classification methods, which is often used as a baseline, is to always predict the most common class label. This is called the **majority classifier**.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Majority Classifier

The most common class label (9/14) is "Yes," so always answer "Yes" no matter what the weather is like.

$$accuracy = \frac{correct}{total}$$

This classifier is 64% accurate.

Naïve Bayes

Now let's explore another simple classifier: a Bayes Net that assumes all evidence variables are conditionally independent.

To predict the class label, we calculate the probability distribution of the query given the evidence and return the most likely value.

This technique scales well and is simple to calculate, so it is also commonly used as a baseline.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	[illegible]	[illegible]	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

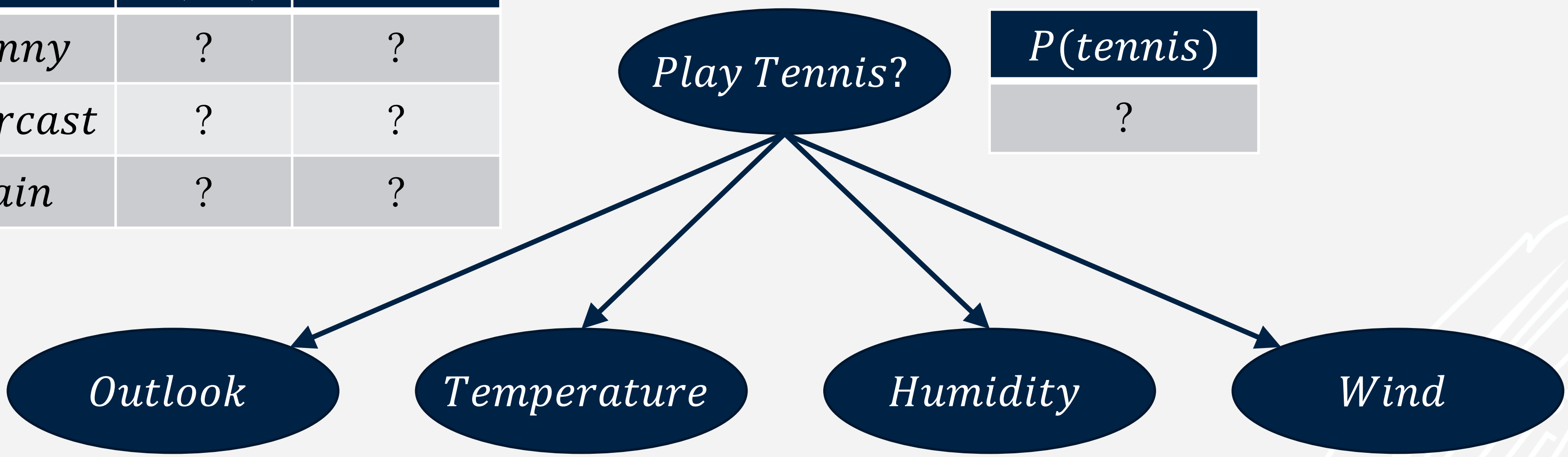
Training Set

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	[illegible]	[illegible]	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Test Set

O	$P(O\|t)$	$P(O\|\neg t)$
sunny	?	?
overcast	?	?
rain	?	?

$P(tennis)$
?



M	$P(M\|t)$	$P(M\|\neg t)$
hot	?	?
mild	?	?
cool	?	?

H	$P(H\|t)$	$P(H\|\neg t)$
high	?	?
normal	?	?

W	$P(W\|t)$	$P(W\|\neg t)$
weak	?	?
strong	?	?

$P(sun|tennis) = ?$

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
(Test Set Hidden)					
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$P(sun|tennis) = \frac{?}{8}$$

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
(Test Set Hidden)					
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$P(sun|tennis) = \frac{2}{8}$$

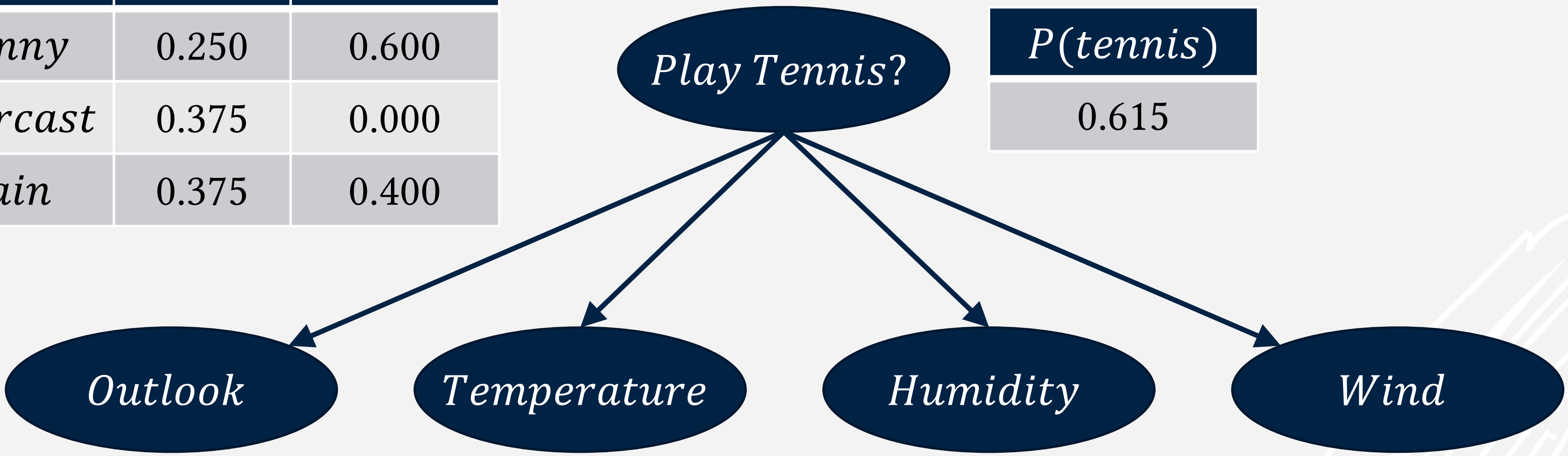
Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
(Test Set Hidden)					
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

NIL

$P(sun|tennis) = .25$

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
(Test Set Hidden)					
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

O	$P(O\|t)$	$P(O\|\neg t)$
sunny	0.250	0.600
overcast	0.375	0.000
rain	0.375	0.400



$P(tennis)$
0.615

M	$P(M\|t)$	$P(M\|\neg t)$
hot	0.125	0.400
mild	0.500	0.400
cool	0.375	0.200

H	$P(H\|t)$	$P(H\|\neg t)$
high	0.250	0.800
normal	0.750	0.200

W	$P(W\|t)$	$P(W\|\neg t)$
weak	0.625	0.400
strong	0.375	0.600

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) = ?$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) =$$
$$\mathbf{P}(Tennis, overcast, hot, high, weak)$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) =$$
$$\mathbf{P}(Tennis)\mathbf{P}(overcast|Tennis)\mathbf{P}(hot|Tennis)$$
$$\mathbf{P}(high|Tennis)\mathbf{P}(weak|Tennis)$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) =$$
$$\alpha\langle 0.615{,}0.385\rangle\langle 0.375{,}0.000\rangle\langle 0.125{,}0.400\rangle$$
$$\langle 0.250{,}0.800\rangle\langle 0.625{,}0.400\rangle$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) = \alpha\langle 0.0045, 0.0000\rangle$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) = \langle 1,0 \rangle$$

Naïve Bayes Classifier

Given the Bayes Net learned on the training set, does it correctly classify the test set?

D3	Overcast	Hot	High	Weak	Yes

$$\mathbf{P}(Tennis|overcast, hot, high, weak) = \langle 1{,}0 \rangle$$

The network predicts that there is a 100% chance that the answer is "yes," and it is correct.

Noise

Sometimes input data has incorrect values called **noise**. Sometimes this can be detected and corrected.

For example, we might accidentally write down "Yes" on a day we did not play.

Overfitting

Sometimes a model learns patterns that appear in the training set but that do not exist in the actual problem. This is called **overfitting**.

For example, the Naïve Bayes classifier will always return "Yes" when it is overcast because in all our examples we played on overcast days. However, there may be overcast days when we do not play.

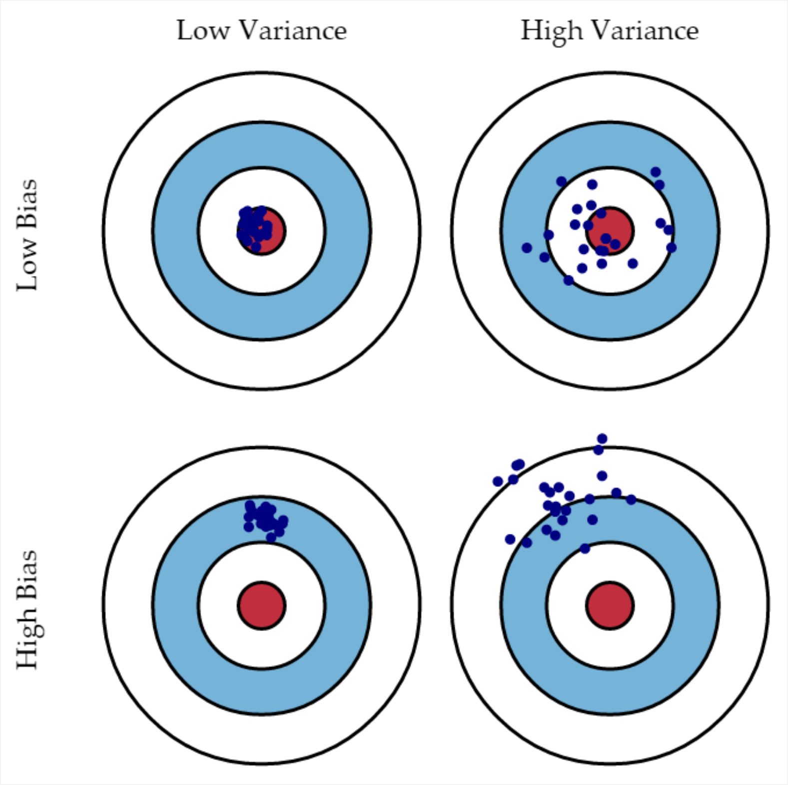
Bias Variance Tradeoff

A classifier's **bias** is error from underfitting. It occurs when a classifier misses some important associations.

A classifier's **variance** is error from overfitting. It occurs when it is too sensitive to small patterns in the training sets that don't exist in the problem.

It is usually impossible to minimize both kinds of error; instead, they form a tradeoff. Many classifiers allow the author to specify bias/variance constraints.

Bias vs. Variance



Source: Scott Fortmann-Roe, "Understanding the Bias Variance Tradeoff."
http://scott.fortmann-roe.com/docs/BiasVariance.html

No Free Lunch

The **No Free Lunch Theorem** states that every classifier's accuracy is equal when averaged over all possible data sets.

In other words, there is no such thing as a "best" classifier in general. The classifier you should use depends on your data.