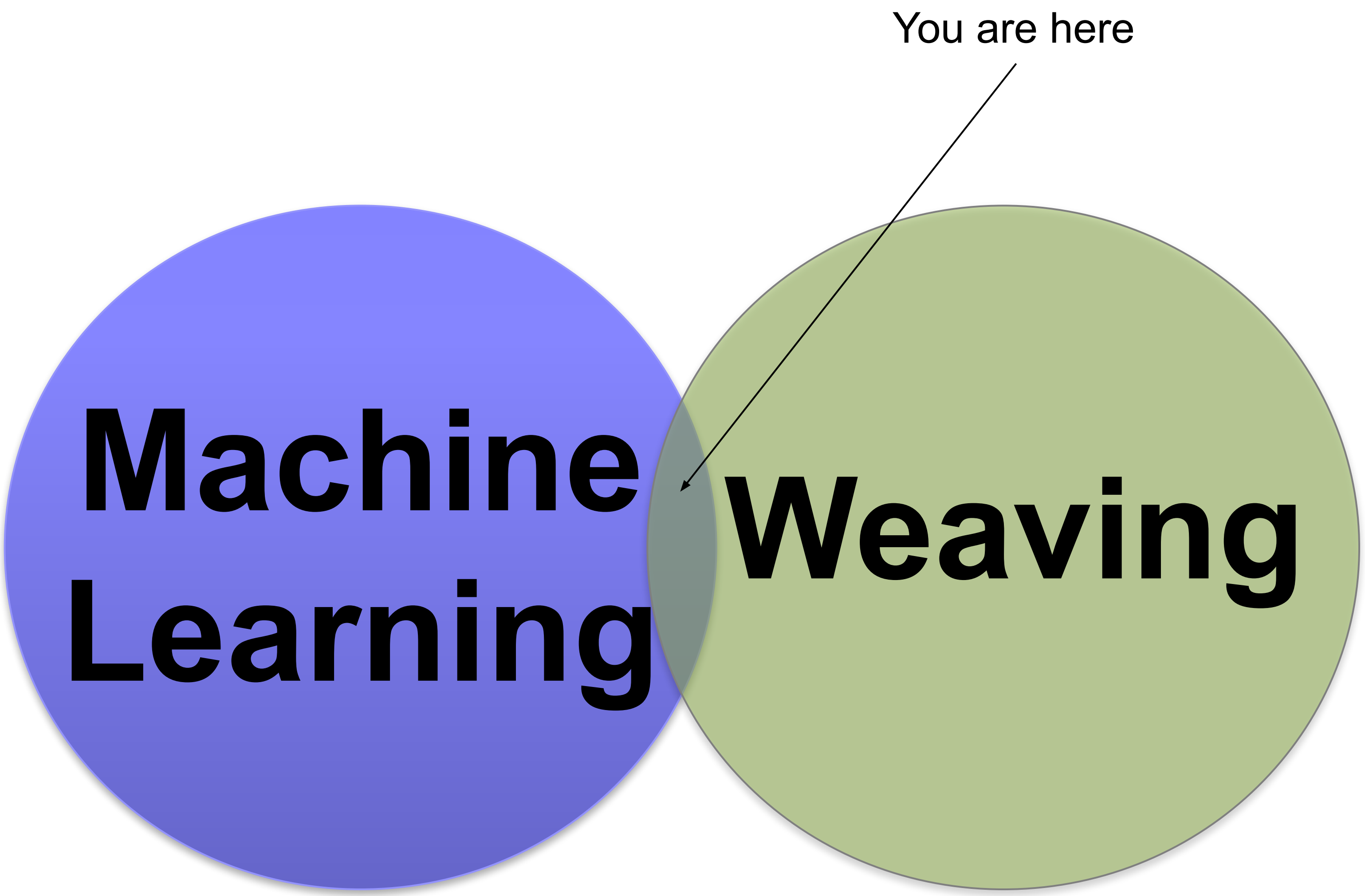
OurCS: Exploring Machine Learning and Automation through Smart Textiles

Carolyn Penstein Rosé and Jim McCann
Language Technologies Institute
Human-Computer Interaction Institute
Robotics Institute

https://tinyurl.com/cmuweave

Intro
9:30-10:30


You are here
Machine Learning
Weaving

What is a pattern?
Tapestry Weaving

What is learning?
Multi-shaft Weaving

Where can automation take us?
Jacquard Weaving

Workshop Schedule

- Fri 9:30-10:30 Introduction to Machine Learning and Weaving, Getting Acquainted
- Fri 10:45-12:30 Distributions, Model Fitting, Tapestry Weaving
- Fri 3-5 Hands on design and weaving instruction
- Fri 6:30-7:30 Hang out and weave
- Sat 10:30-12:30 Introduction to Matrix Factorization and Pattern Weave
- Sat 4pm-6pm Weaving Software Exploration, Machine Learning and Textiles
- Sun 9-10:30 Jacquard Weaving
- Sun 10:30-12 Prepare presentations

Who am I?

PhD in Language and Information Technologies, 1998

Enjoys Israeli folk dancing, playing piano, long walks in the woods, knitting, crocheting, weaving, and spinning yarn

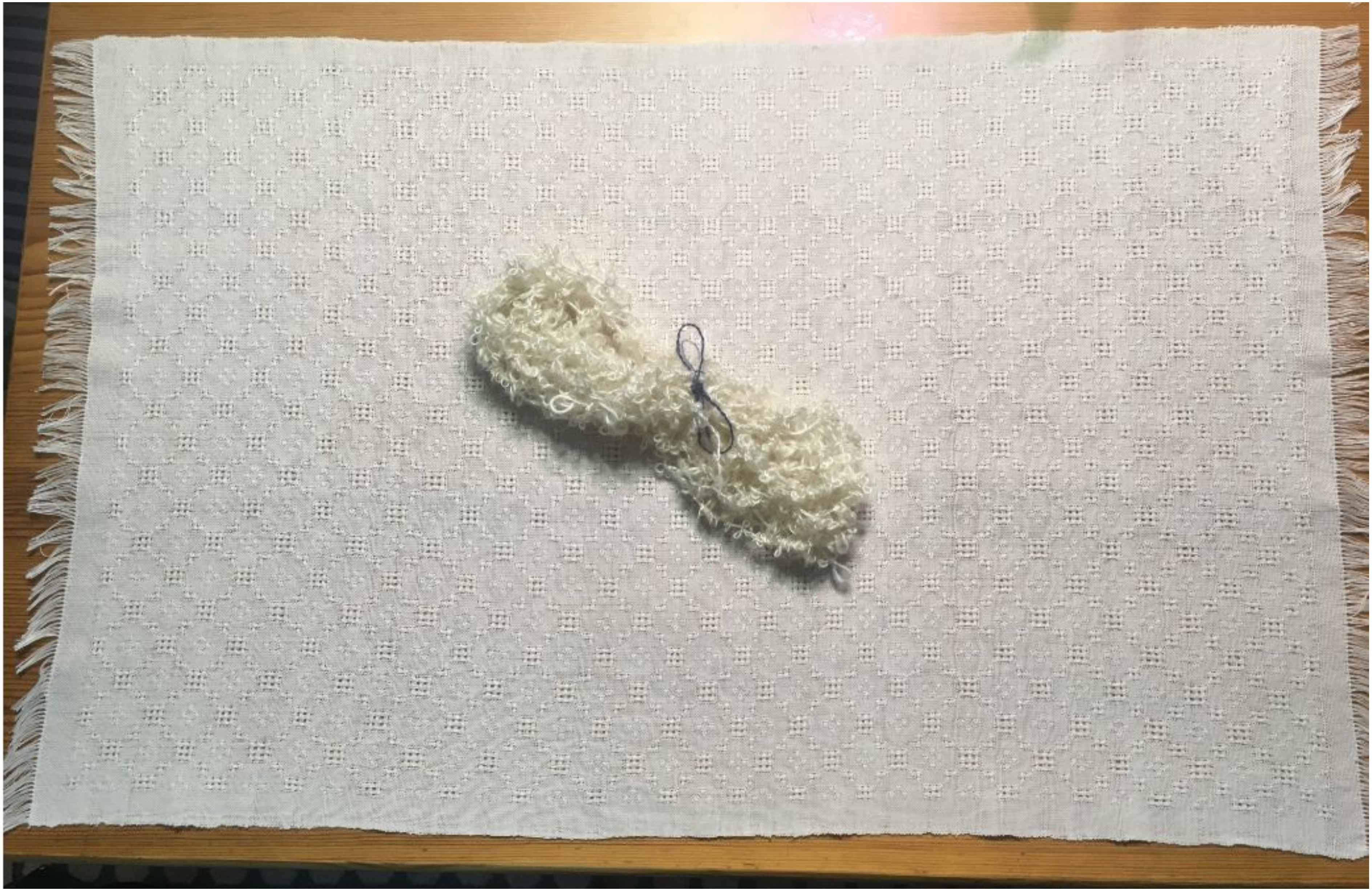
- Joint appointment between Language Technologies and Human-Computer Interaction
- Research in Computational Discourse Analysis and Computer-Supported Collaborative Learning
- Past President and Inaugural Fellow of the International Society of the Learning Sciences
- Executive Editor of the International Journal of Computer Supported Collaborative Learning

Who is Jim?

PhD in Computer Science, 2010

Also, as time allows, makes video games as "TCHOW llc" (see tchow.com) and music as "Jimike" (see Spotify)

- Assistant Professor in the Robotics Institute
- Research in making tools to enable "Universal Creativity"
- Textiles Lab looks at improving the state of the art in textiles design and fabrication.
- General interest in automation that helps improve human experience



Tell us about you!

3 truths and a lie

Introduction to Machine Learning and Weaving 10:45-11:30

Overwhelmed with data…

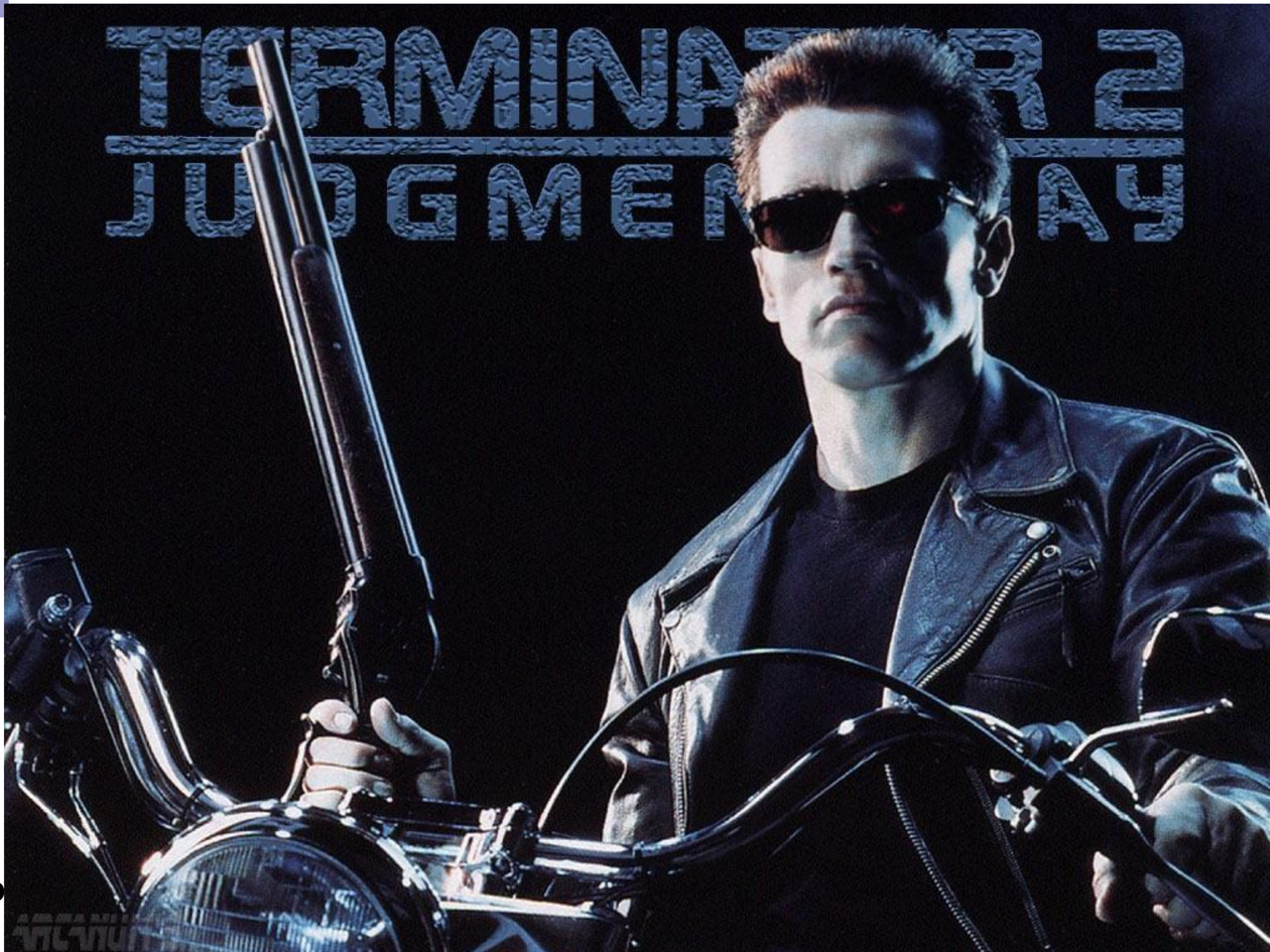
www.powerfulinformation.org

What do we do with all of it?

Machine learning is about automatically finding meaningful patterns in data

Example for credit history data:

Rule predicts who is more likely to have problems paying off credit.

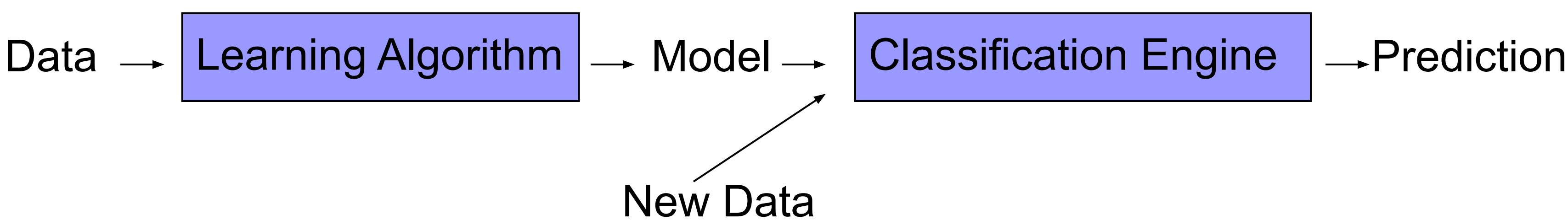


Ap

an

What is machine learning?

- Automatically or *semi-automatically*
  - Inducing concepts (i.e., rules) from data
  - Finding patterns in data
  - Explaining data
  - Making predictions



What will be the prediction?

Model

Outlook:
- Sunny -> No
- Overcast -> Yes
- Rainy-> Yes

New Data

outlook	temperature	humidity	windy	play
rainy	cool	high	FALSE	Yes

How does machine learning work?

outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunn			E	yes
rainy			E	yes
s				yes
c				yes
overcast			ALSE	yes
rainy			TRUE	no

A slightly more sophisticated rule learner will find t… the most informat… ass. What do you t… his case?

Outlook:

- Sunny -> No
- Overcast -> Yes
- Rainy-> Yes

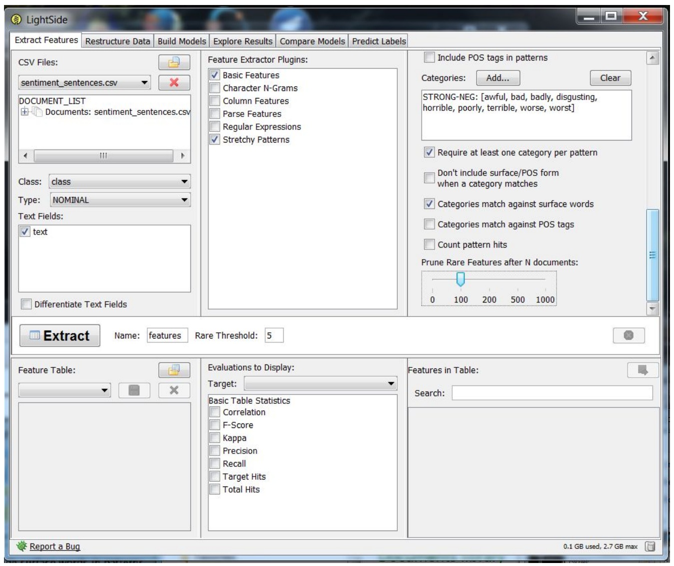
Clas…

What w… se?

<Feature Name>:

- <value> -> <prediction>
- <value> -> <prediction>
- …

LightSIDE

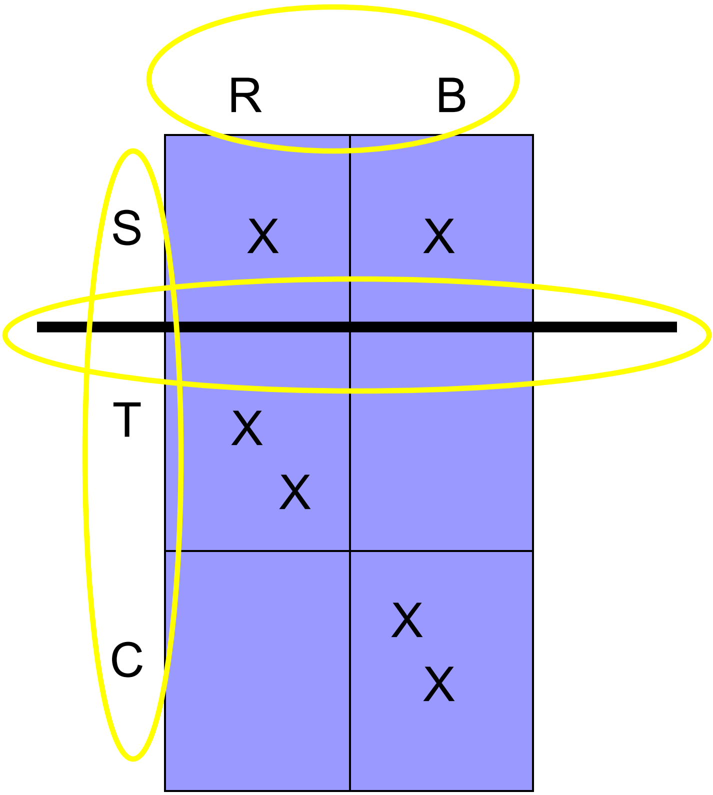


- Download the tool: http://ankara.lti.cs.cmu.edu/side/download.html
- Watch YouTube videos: https://www.youtube.com/playlist?list=PL_ICWHSmCge20_G3qzAF5A1wpa7VxYi7F

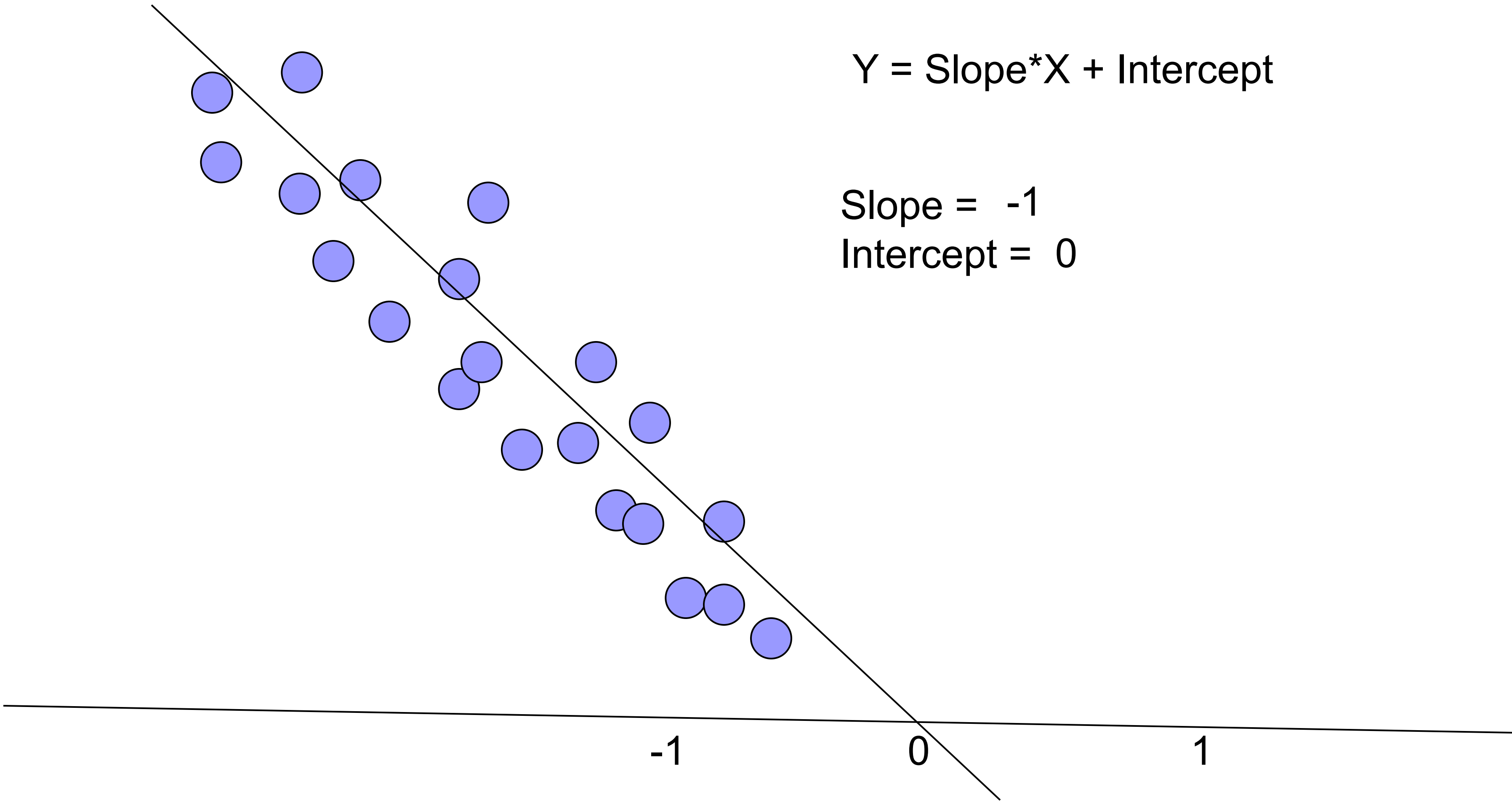
Distributions, Model Fitting, and Tapestry Weaving 11:30-12:30

Patterns as Lines

Color	Shape	Cost
Red	Square	Expensive
Red	Triangle	Cheap
Blue	Circle	Cheap
Blue	Circle	Cheap
Red	Triangle	Cheap
Blue	Square	Expensive



What does it mean to fit a function to some data points?



Y = Slope*X + Intercept

Slope = -1
Intercept = 0

Sample Problem

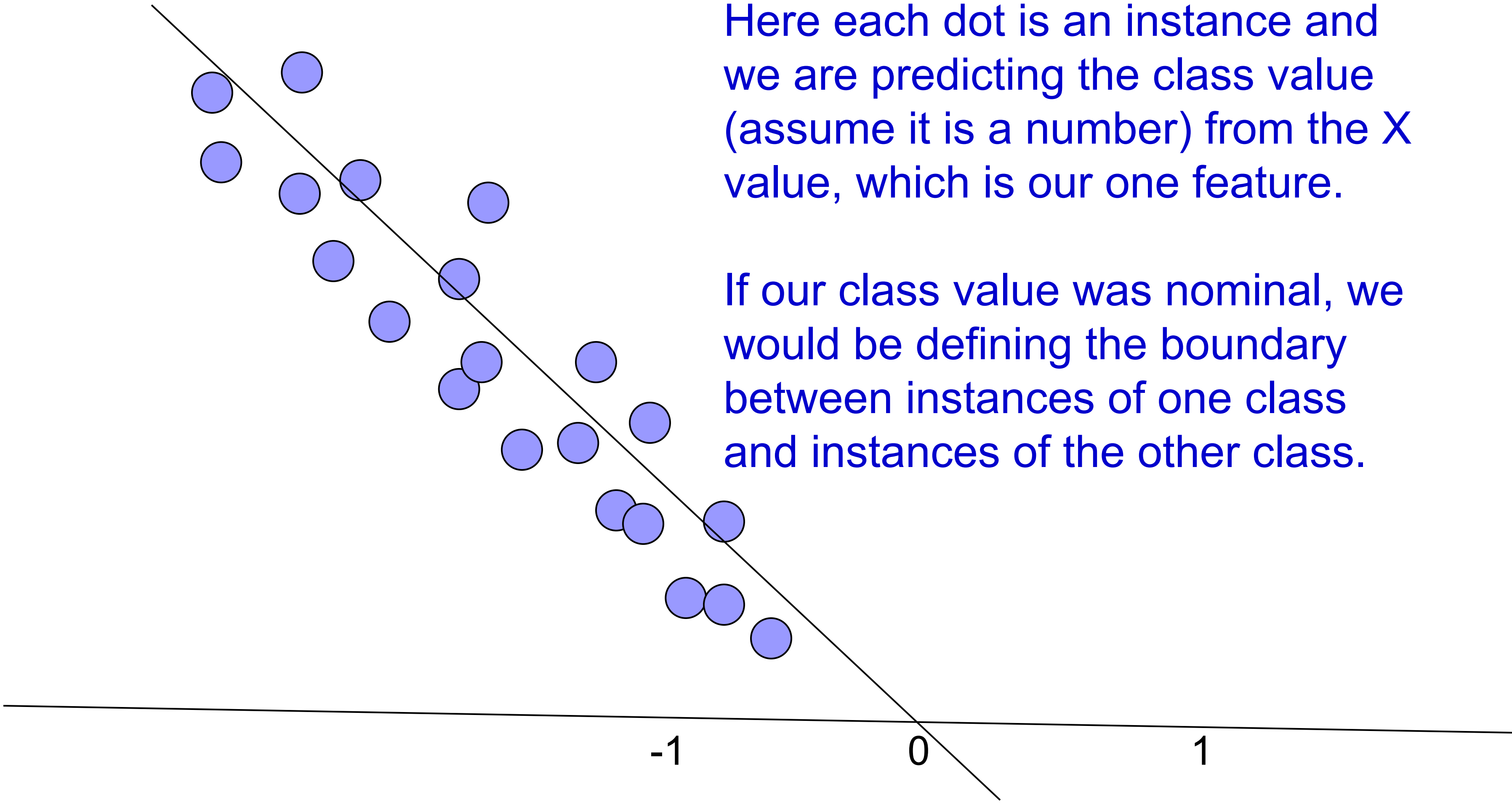
A	B	C
Temperature	Humidity	Play
65.63854838	75.19501097	yes
81.72583232	58.75378279	yes
87.94397182	64.93042496	yes
73.32638942	89.23827594	yes
95.61201149	93.4410086	no
83.03217237	59.87646875	yes
50.39678536	96.83678291	yes
99.13162741	79.11466973	no
62.52431284	98.28645934	yes
90.45078477	55.24912766	yes
82.45952477	81.97602596	no
58.95385441	55.32946544	yes
54.13257904	95.95167081	yes
98.29408575	72.69458699	yes
91.26965556	59.7400881	yes
93.41380703	75.8435062	no
60.38478895	78.28273456	yes
69.33236269	88.41118466	yes
57.53443779	95.14254048	yes
70.37578476	88.57145324	yes

If (Temp > 75) and (Humidity > 75)
  Then No
  Otherwise Yes

Best Linear Model:
Y=.29*Temp + .23*Hum – 43
If Y > 0, then No otherwise Yes

Think about what the coefficients mean!

What does it mean to fit a function to some data points?

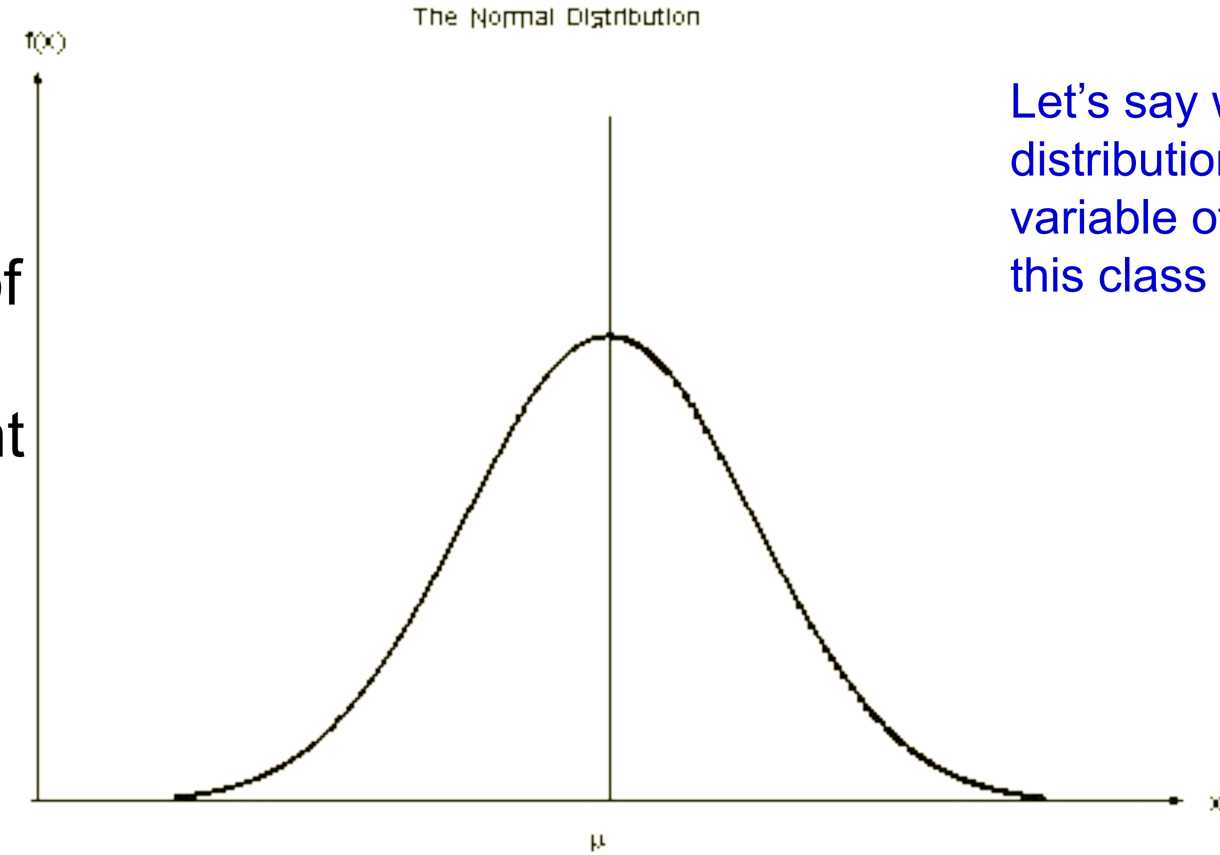


Here each dot is an instance and we are predicting the class value (assume it is a number) from the X value, which is our one feature.

If our class value was nominal, we would be defining the boundary between instances of one class and instances of the other class.

Another kind of function we try to fit is a histogram of Y values, which we call a Probability Distribution Function.

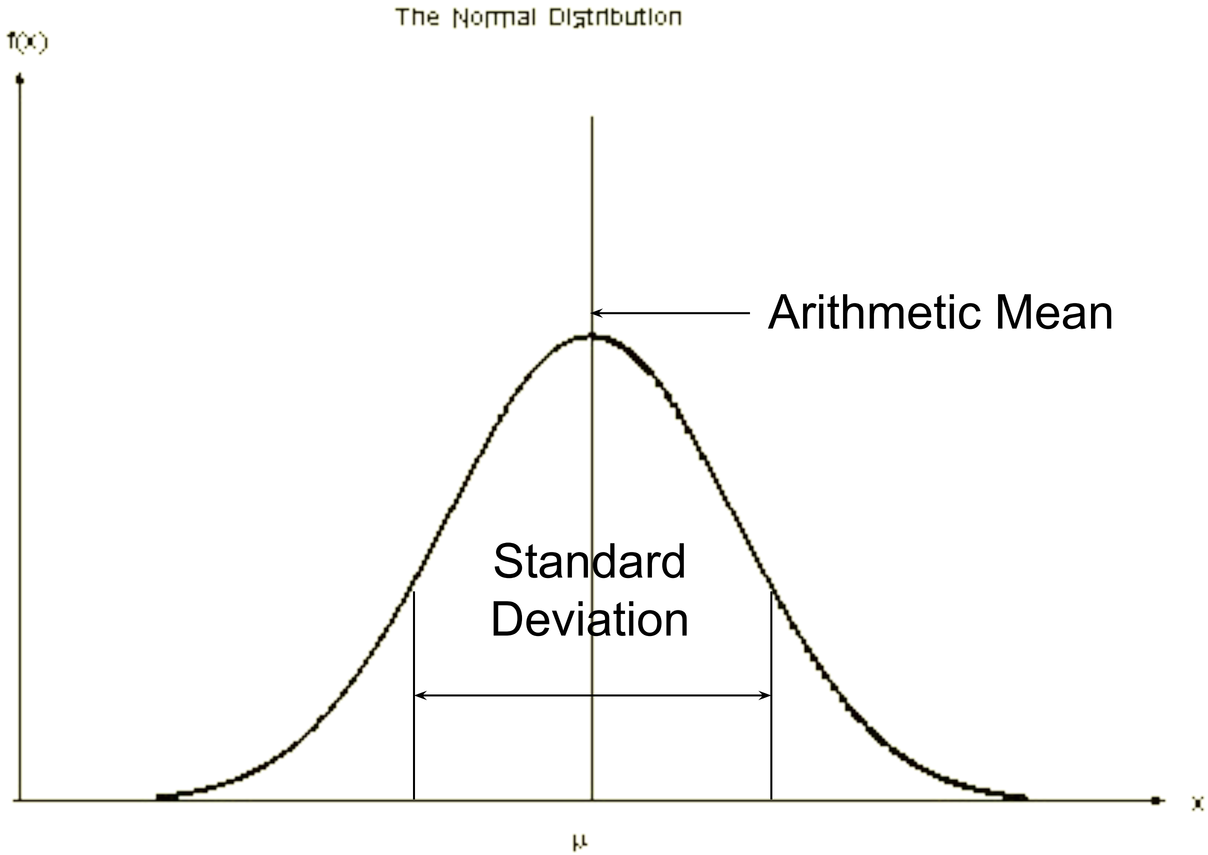
Number of students with height



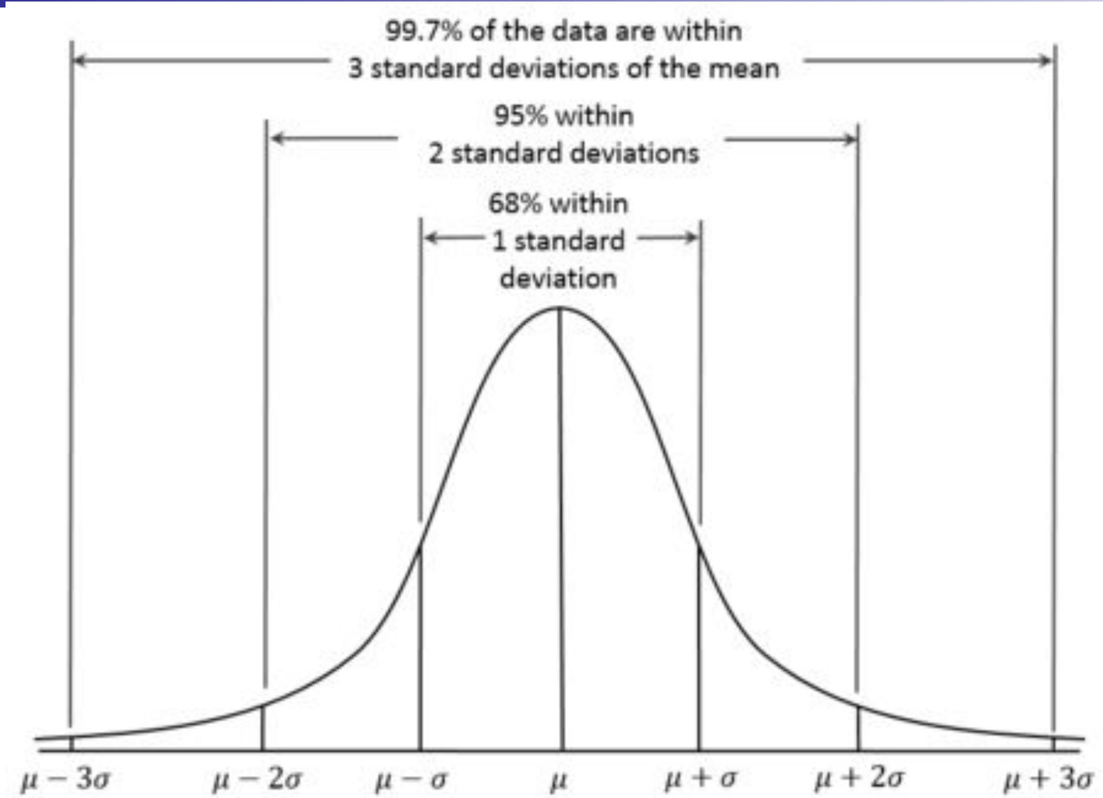
Height

Let’s say we want a distribution of a height variable of students in this class

What does it mean to fit a distribution to some data points?

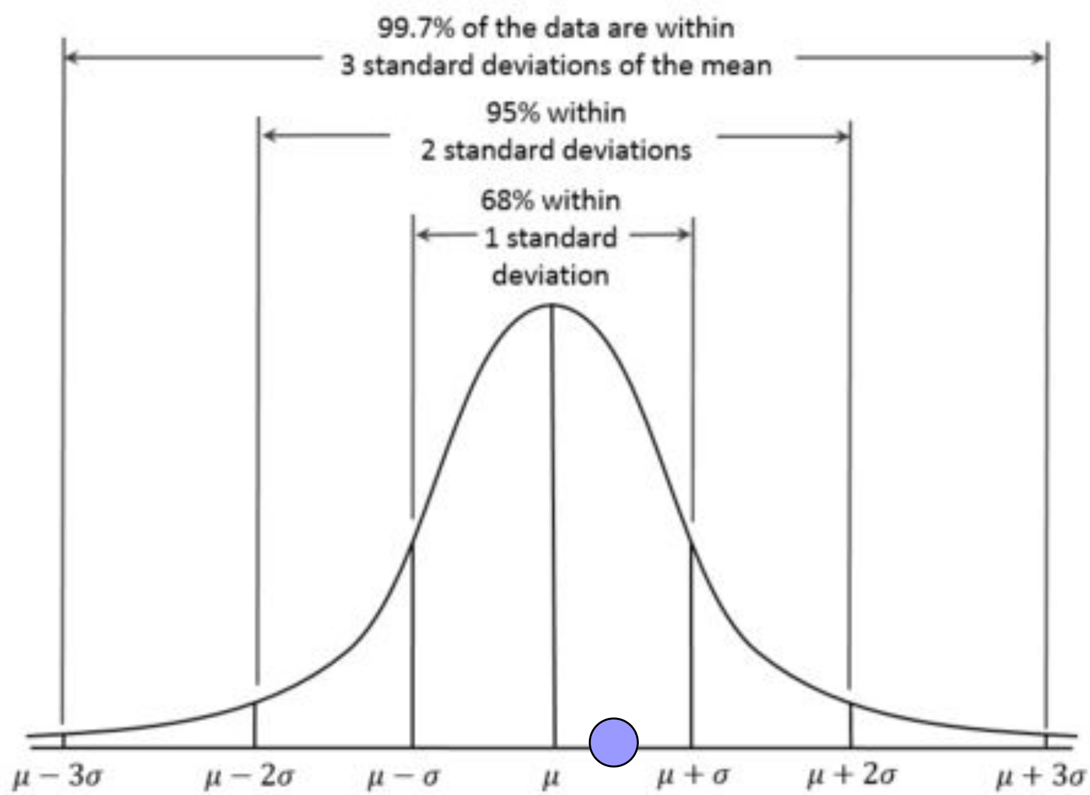


… And from this you can compute probabilities of specific x positions!



Define a binary variable that asks whether a value is within a particular probability band

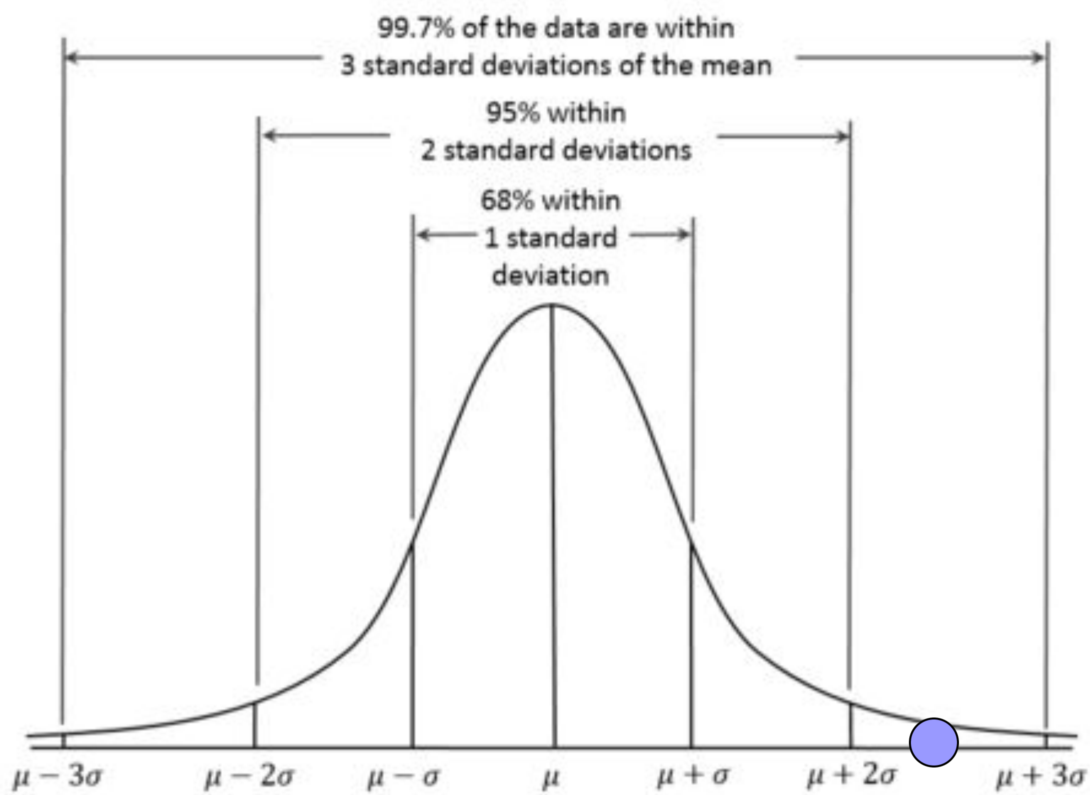
VarA is true if the value is within one standard deviation
VabB is true if the value is within two standard deviations



Define a binary variable that asks whether a value is within a particular probability band

VarA is true if the value is within one standard deviation
VabB is true if the value is within two standard deviations

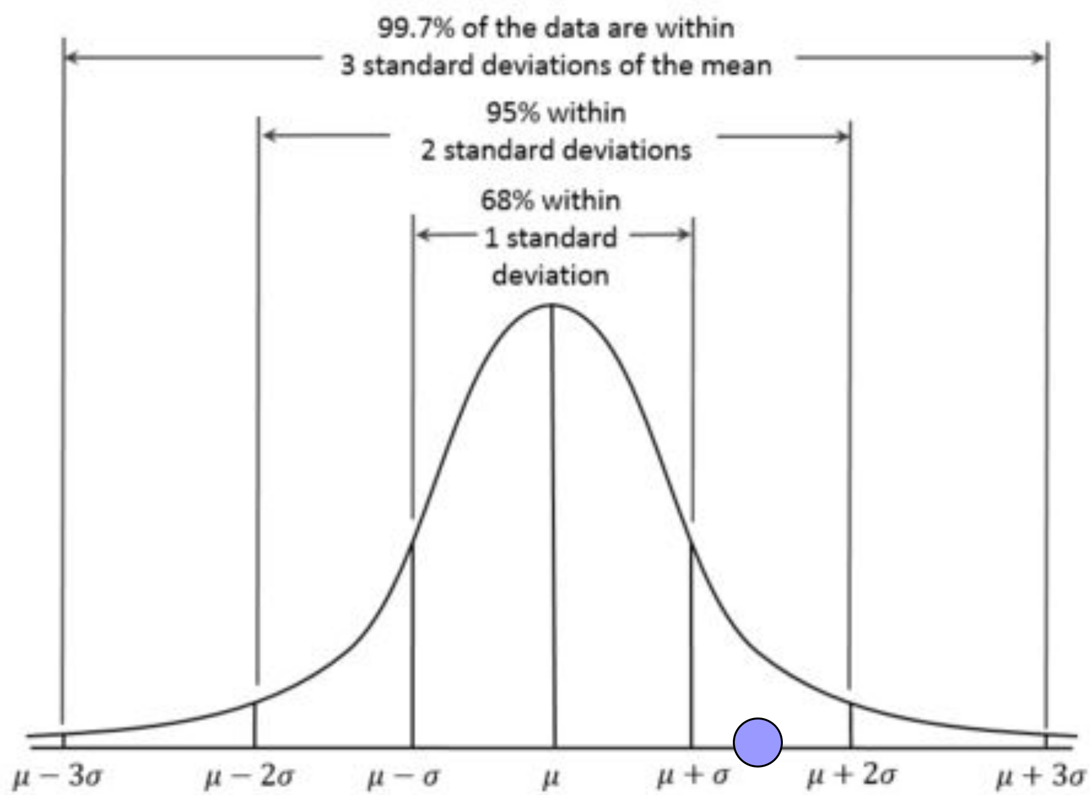
VarA = True
VarB= True



Define a binary variable that asks whether a value is within a particular probability band

VarA is true if the value is within one standard deviation
VabB is true if the value is within two standard deviations

VarA = False
VarB= False



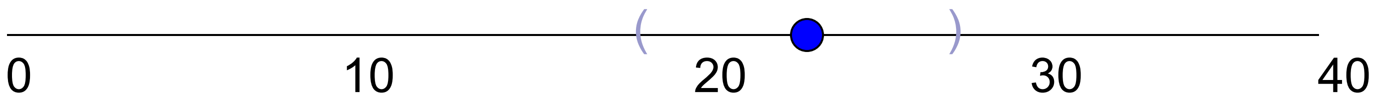
Define a binary variable that asks whether a value is within a particular probability band

VarA is true if the value is within one standard deviation
VabB is true if the value is within two standard deviations

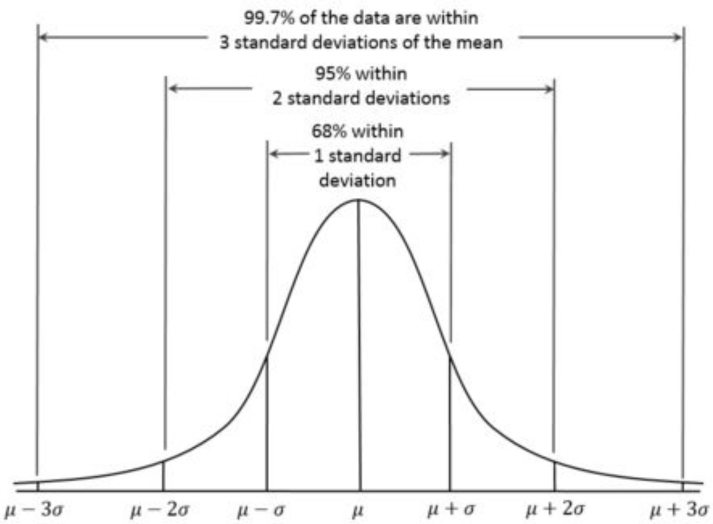
VarA = False
VarB= True

Confidence Intervals

- We know that the accuracy measured on each fold of a cross validation will be different
- The average across these is a better estimate of expected performance than any of the separate ones
- Confidence intervals allow us to say that the probability of the real performance value being within a certain range from the observed average value is within a confidence level – like 90%

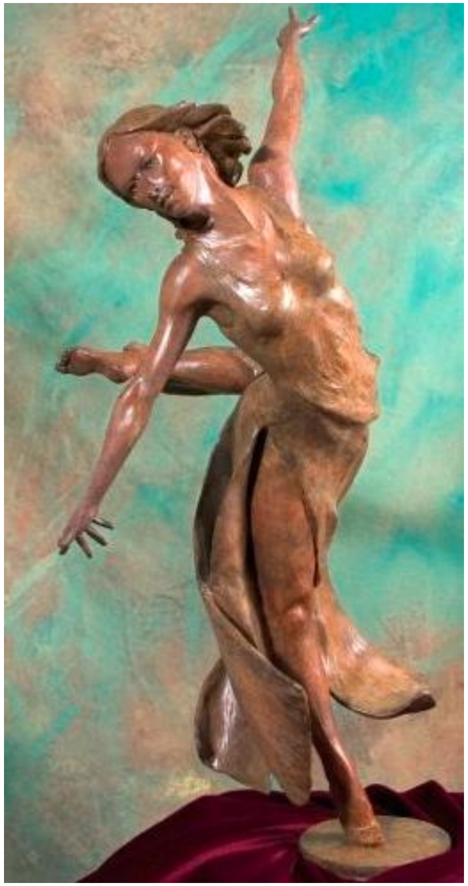


Confidence Intervals



- Confidence limits come from the normal distribution
- Computed in terms of number of standard deviations from the mean
- If the data is normally distributed, there is a 15% chance of the real value being more than 1 standard deviation above the mean

What is a pattern?



“Gyre” by Eric Rosé

Low resolution rendition

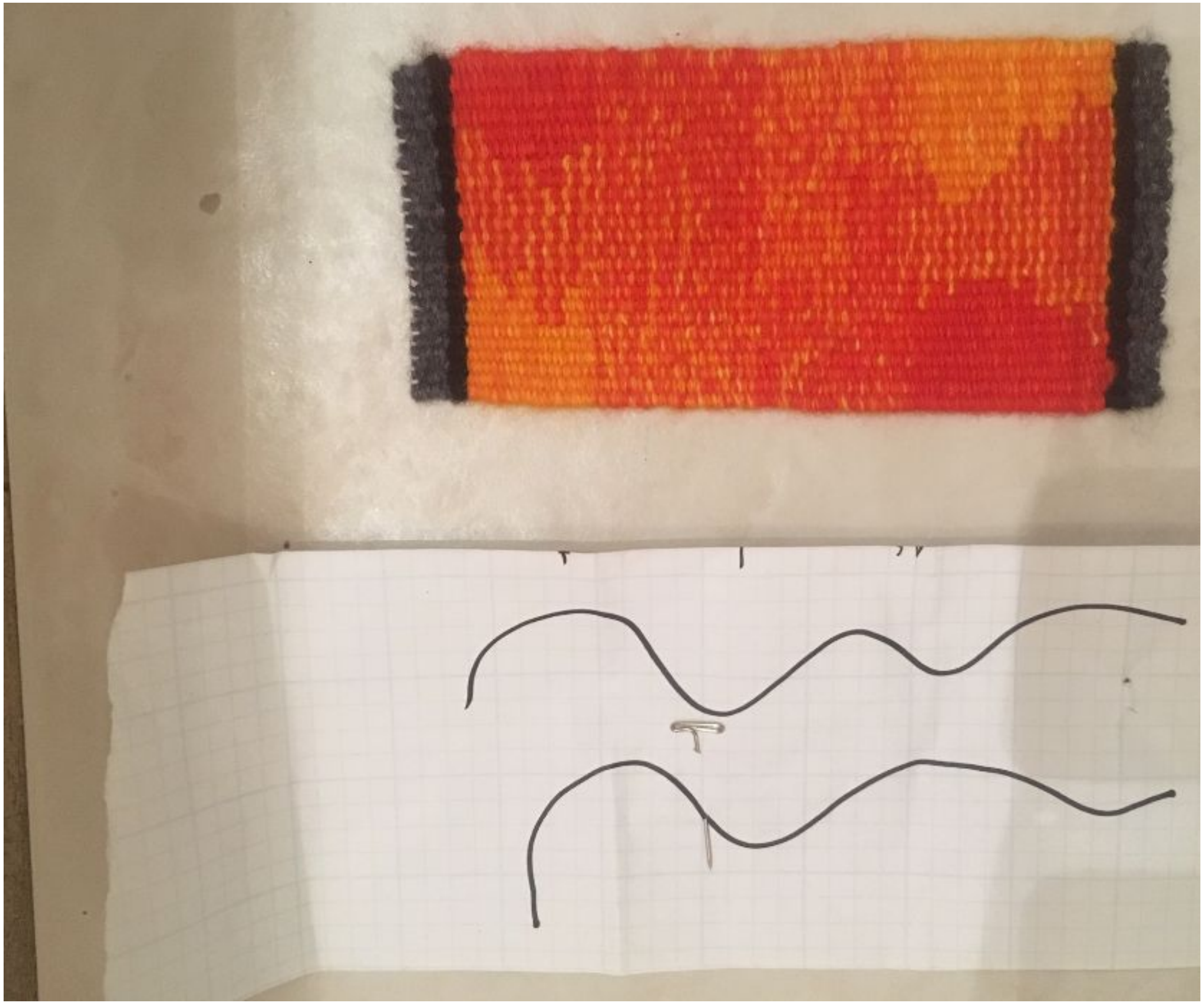
Same pattern, higher resolution

Noise causes interference in detection of the pattern

Low resolution gives more information if there is less interference

What if there are
multiple kinds of patterns
happening at the same time?

Pattern Interference



Pattern Interference

What is different in how men and women talk?

- Word-based features capture more content than style, and are thus vulnerable to domain specificity.

male	female
linux	shopping
microsoft	mom
gaming	cried
server	freaked
software	pink

(Schler 2006)

What is different in how men and women talk?

- Women's language as "deviant" - *Lakoff (1975)* or "more varied" - *Chambers (1992)*

- Extrathematic details in conversational storytelling - time and location (male), people and speech acts (female). *Johnstone (1993)*
  "...after a full three years..."
  "...he would sit and talk to my mother..."

What is different in how men and women talk?

- Hedging, qualifiers, and intensifiers -
  "I think I might have said ..."
  "So he brought to me..."
  "I'm sometimes so jealous of people"

- "like" particle - gender variations in placement and usage
  Iyeiri, Yaguchi, Okabe (2005)
  *"...and then, we asked **like** four and one..."*
  *"**Like**, instead of advanced, basic, proficient, and whatever..."*

Confounded with other variables

- Men sound older and women sound younger (Argamon et al., 2007)
- Men sound more like non-fiction and women sound more like fiction (Argamon et al., 2003)

Pattern Interference

Train

FYH MYL
MYH MOL
MYL
MYL
MYL
MYH MOL
FOH
MYL
MOL
MYH MYL
MOL
MYL
MOL
MYL
MYH
MOL FYH
MYL
MYH
FYH
MOL FYH
FYH
MYH FYH
MYL
MOL
MYH
FOH
FOH
FOH
MYH
FOH
FOH
FYH
FOH
FOH

Test

FYL MOH
MOH
FOL MOH
MOH
FYL FOL
FOL
FYL
MOL
FOL
FYL
MOL
MOH
FOL
MOH
MYH
MOL
FYL
MOH
MYH
MYH MOH
FYL
MYH
FYL
MOH

Sample Problem (Continued)

A	B	C
Temperature	Humidity	Play
65.63854838	75.19501097	yes
81.72583232	58.75378279	yes
87.94397182	64.93042496	yes
73.32638942	89.23827594	yes
95.61201149	93.4410086	no
83.03217237	59.87646875	yes
50.39678536	96.83678291	yes
99.13162741	79.11466973	no
62.52431284	98.28645934	yes
90.45078477	55.24912766	yes
82.45952477	81.97602596	no
58.95385441	55.32946544	yes
54.13257904	95.95167081	yes
98.29408575	72.69458699	yes
91.26965556	59.7400881	yes
93.41380703	75.8435062	no
60.38478895	78.28273456	yes
69.33236269	88.41118466	yes
57.53443779	95.14254048	yes
70.37578476	88.57145324	yes

If (Temp > 75) and (Humidity > 75)
Then No
Otherwise Yes

Best Linear Model:
Y=.29*Temp + .23*Hum – 43
If Y > 0, then No otherwise Yes

Error Analysis Process

High Level Overview

```
=== Confusion Matrix ===

  a   b   <-- classified as
 44  25 |
 10 149 |
```

Goal: We want to discover how to re-represent the data so that instances with the same class value look more similar to one another and instances with different class values look more different

- Identify large error cells
- Make comparisons
  - Ask yourself how it is similar to the instances that were correctly classified with the same class (**vertical comparison**)
  - How it is different from those it was incorrectly not classified as (**horizontal comparison**)

Doing an Error Analysis

Temperatı	Humidity	Combine	Play	Predicted	AveTemp	AveHum
98.01125	77.75734	1	no	no	93.76555	87.43199
94.78822	90.76342	1	no	no		
88.81202	82.3188	1	no	no		
95.99637	87.94525	1	no	no		
94.09122	94.419	1	no	no		
90.89421	91.38811	1	no	no		
75.66254	94.21484	1	no	no		
78.1215	84.59609	1	no	yes	77.76203	82.7527
77.40257	80.90931	1	no	yes		
70.58709	97.96222	0	yes	no	72.2215	95.99596
73.85592	94.0297	0	yes	no		
65.90603	68.24662	0	yes	yes	71.33624	70.51335
97.56169	57.61119	0	yes	yes		
66.02028	98.86424	0	yes	yes		
69.82286	52.19364	0	yes	yes		
62.10335	99.23564	0	yes	yes		
68.93071	82.11282	0	yes	yes		
61.23483	75.94503	0	yes	yes		
56.58624	50.70179	0	yes	yes		
99.94828	59.30695	0	yes	yes		

	Yes	No
Yes	18	2
No	2	7

"Y=.29*Temp + .23*Hum - 43"

	hor-temp	hor-hum	ver-temp	ver-hum
yes-no	0.885259	**25.48261**	21.54405	8.56397
no-yes	**16.00351**	8.56397	6.425794	12.23935

The problem is that the linear model can't see
the connection between Temperature and Humidity.

Doing an Error Analysis

Temperat	Humidity	Combine	Play	Predicted		AveTemp	AveHum
98.01125	77.75734	1	no	no		93.76555	87.43199
94.78822	90.76342	1	no	no			
88.81202	82.3188	1	no	no			
95.99637	87.94525	1	no	no			
94.09122	94.419	1	no	no			
90.89421	91.38811	1	no	no			
75.66254	94.21484	1	no	no			
78.1215	84.59609	1	no	yes		77.76203	82.7527
77.40257	80.90931	1	no	yes			
70.58709	97.96222	0	yes	no		72.2215	95.99596
73.85592	94.0297	0	yes	no			
65.90603	68.24662	0	yes	yes		71.33624	70.51335
97.56169	57.61119	0	yes	yes			
66.02028	98.86424	0	yes	yes			
69.82286	52.19364	0	yes	yes			
62.10335	99.23564	0	yes	yes			
68.93071	82.11282	0	yes	yes			
61.23483	75.94503	0	yes	yes			
56.58624	50.70179	0	yes	yes			
99.94828	59.30695	0	yes	yes			

	Yes	**No**
Yes	18	2
No	2	7

"Y=.29*Temp + .23*Hum - 43"

	hor-temp	hor-hum	ver-temp	ver-hum
yes-no	0.885259	**25.48261**	21.54405	8.56397
no-yes	**16.00351**	8.56397	6.425794	12.23935

Horizontal comparison is high because the high Humidity values make these instances look different from Yes cases. However, there are Yes cases with somewhat high Humitidy values but low Tempeterature values, and these do have low Temperature values. And if you look at the no cases, though these high Humidity values look more like the high Humidity values of the No cases, those come along with High Tempertature values too, and that is missing from the instances that were incorrectly classified. As we use our world knowledge to reason about what might be being missed in the representation, we are reminded that we don't like playing tenis when it is both hot and humid, and we notice when we make these comparisons that the model is not able to reason that way. So we know we need to add some way for heat and humidity to be considered jointly.

Add a new feature

`=IF(AND(A2>75,B2>75),1,0)`

Temperat	Humidity	**Combine**	Play	Predicted	AveTemp	AveHum
98.01125	77.75734	**1**	no	no	93.76555	87.43199
94.78822	90.76342	**1**	no	no		
88.81202	82.3188	**1**	no	no		
95.99637	87.94525	**1**	no	no		
94.09122	94.419	**1**	no	no		
90.89421	91.38811	**1**	no	no		
75.66254	94.21484	**1**	no	no		
78.1215	84.59609	**1**	no	yes	77.76203	82.7527
77.40257	80.90931	**1**	no	yes		
70.58709	97.96222	**0**	yes	no	72.2215	95.99596
73.85592	94.0297	**0**	yes	no		
65.90603	68.24662	**0**	yes	yes	71.33624	70.51335
97.56169	57.61119	**0**	yes	yes		
66.02028	98.86424	**0**	yes	yes		
69.82286	52.19364	**0**	yes	yes		
62.10335	99.23564	**0**	yes	yes		
68.93071	82.11282	**0**	yes	yes		
61.23483	75.94503	**0**	yes	yes		
56.58624	50.70179	**0**	yes	yes		
99.94828	59.30695	**0**	yes	yes		

	Yes	**No**
Yes	18	2
No	2	7

"Y=.29*Temp + .23*Hum - 43"

	hor-temp	hor-hum	ver-temp	ver-hum
yes-no	0.885259	**25.48261**	21.54405	8.56397
no-yes	**16.00351**	8.56397	6.425794	12.23935

Design and Weaving Instruction

Resources

- https://americantapestryalliance.org/tapestry-education/tapestry-weaving-technique-videos/
- https://www.youtube.com/channel/UCHjPnwX7yFjsdzEC_bznjWQ
- https://weaversloft.com/catalog/wlimages/Tapestry_Weaving.pdf
- https://cdn.shopify.com/s/files/1/2375/8991/files/easywarpsamloombeginningtapestry.pdf

Combining Patterns through Color Blended Threads