

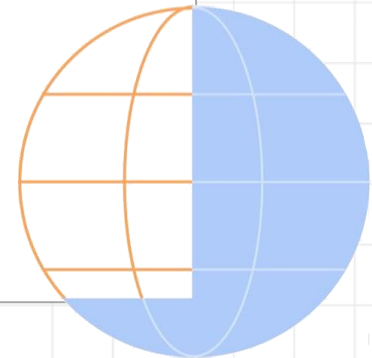


# Day 04

Data Science with Python



Pasindu Marasinghe  
ppm@ucsc.cmb.ac.lk

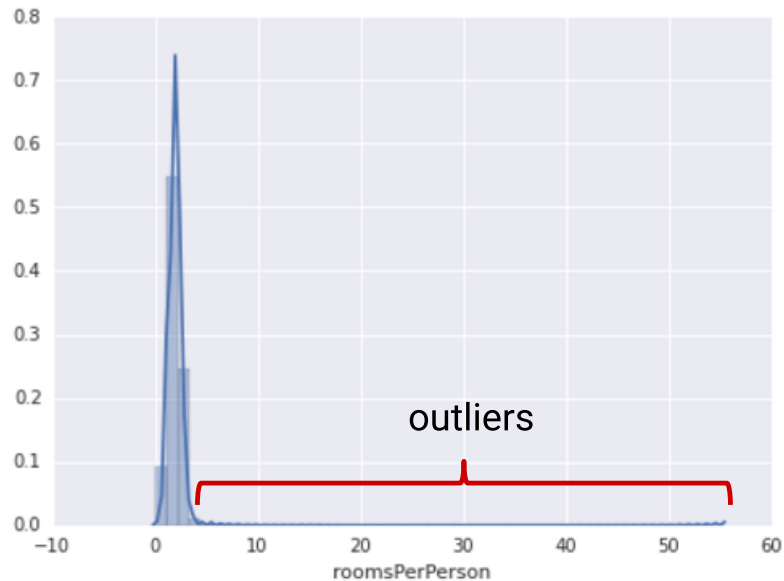


# Why Do We Have to Normalize Data?

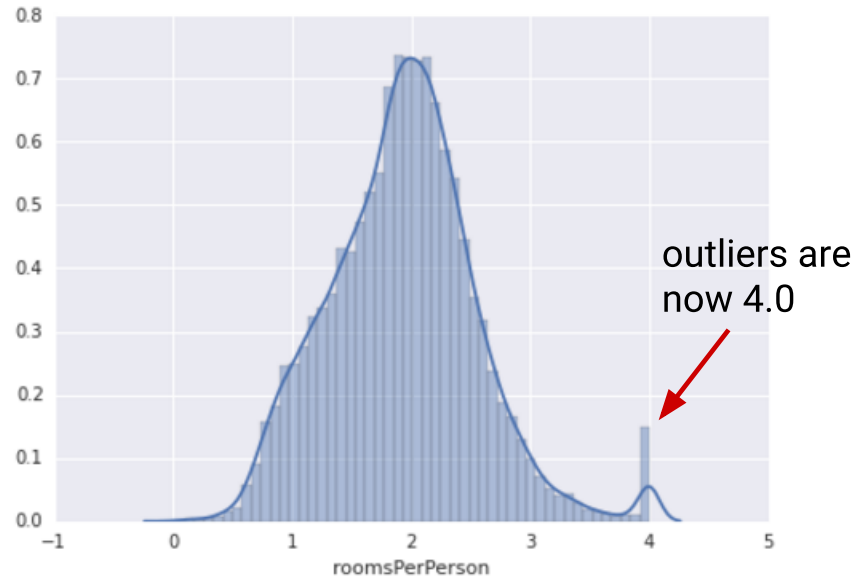
## Scaling to a range

$$X_{new} = \frac{(X - X_{min})}{(X_{max} - X_{min})}$$

# Clipping

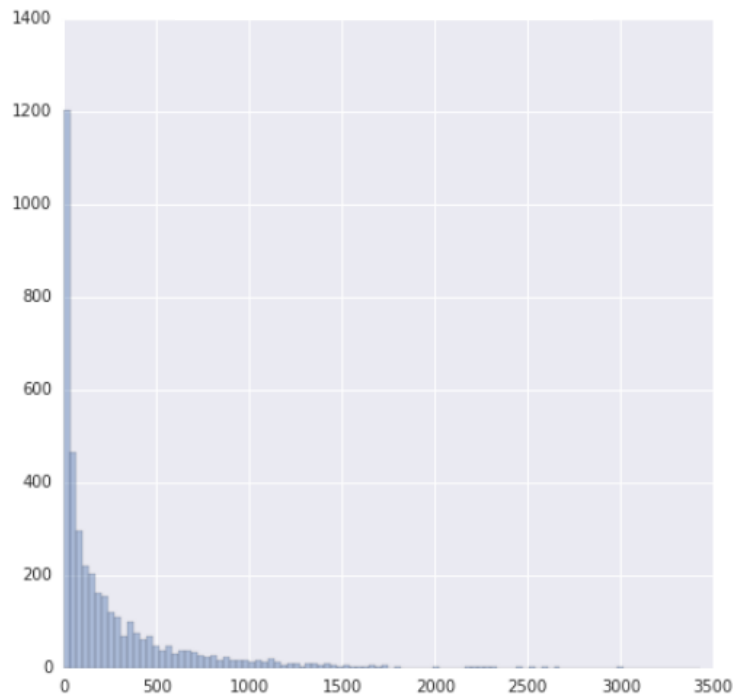


Same feature, capped to a max of 4.0

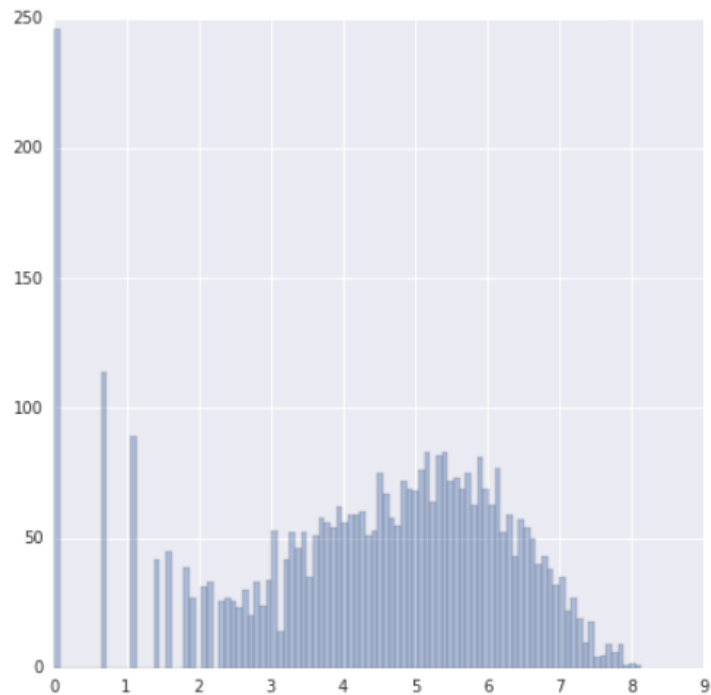


# Log Scaling

Ratings per movie



Log ratings per movie



## Z-score

$$X_{new} = \frac{(X - \mu)}{\sigma}$$

## Regression



What will be the temperature tomorrow?

84°



Fahrenheit

## Classification



Will it be hot or cold tomorrow?

COLD

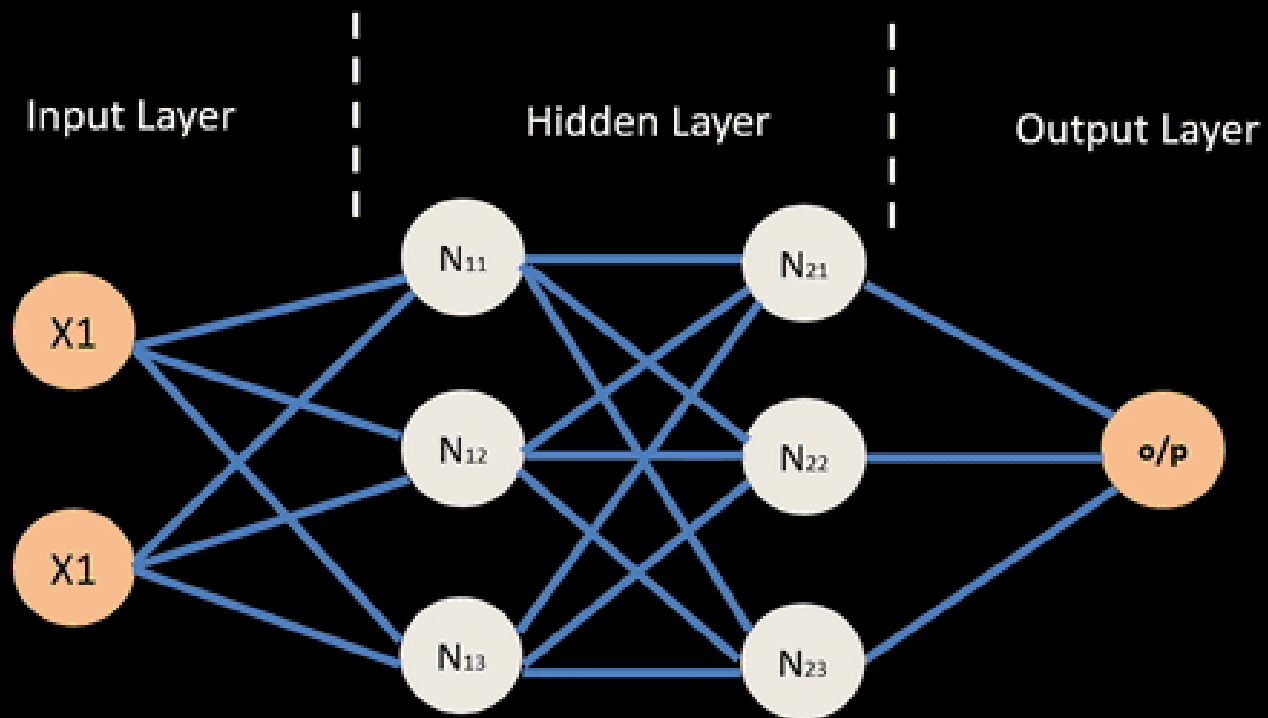
HOT



Fahrenheit

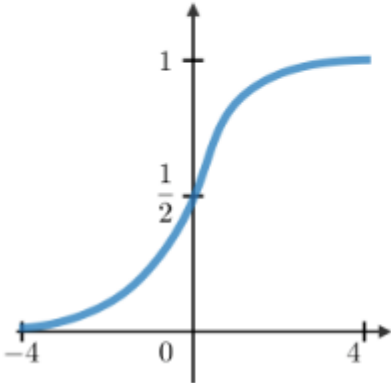
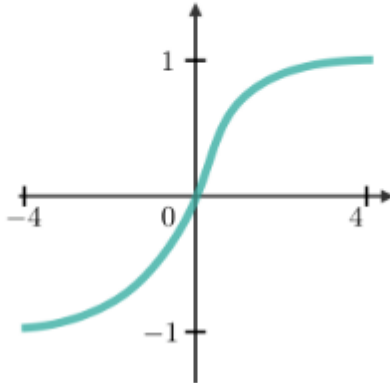
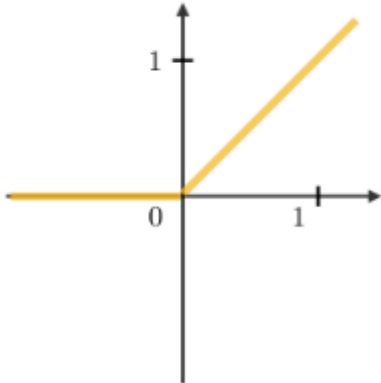
ANN = MLP ?





# Activation Function

# Sigmoid vs Relu vs tanh

Sigmoid	Tanh	RELU
$g(z) = \frac{1}{1 + e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$
		

# Model Evaluation

## Confusion Matrix

		Actual	
		Positive	Negative
Predicted	Positive	<b>True Positive</b>	<b>False Positive</b>
	Negative	<b>False Negative</b>	<b>True Negative</b>

## Accuracy

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

## Precision

$$Precision = \frac{(TP)}{(TP + FP)}$$

## Recall

$$\textit{Recall} = \frac{(TP)}{(TP + FN)}$$



## F1 Score

$$F1\ Score = 2 \frac{(Precision * Recall)}{(Precision + Recall)}$$