Logistic Regression

Recap

WHAT IS A MODEL?

"A representation of some aspect of the world which is based on simplifying assumptions."

Alternatively:

"A phenomenon represented mathematically to produce a simplified version of reality."

REGRESSION

A regression problem is one where the output variable is a real or continuous value, such as "salary" or "weight".

CLASSIFICATION

A classification problem is when the output variable is a category, such as "red" or "blue" / "disease" or "no disease".

WHAT IS LINEAR REGRESSION?

Linear Regression is a way to predict the relationship between two variables, where the dependent variable is continuous.

- Single-variable regression: one predictor, one response

- Multi-variable regression: multiple predictors, one response

Logistic Regression

Umm, so, we'll actually be using it for classification

🤯🤷‍♀️🤯

WHAT IS LOGISTIC REGRESSION?

Logistic regression is a classification algorithm that predicts the probabilities of a categorical (or binary) dependent variable.

WHAT IS LOGISTIC REGRESSION?

Given existing data, it predicts the class/categories for new data

Logistic regression is a classification algorithm that predicts the probabilities of a categorical (or binary) dependent variable.

WHAT IS LOGISTIC REGRESSION?

Logistic regression is a classification algorithm that predicts the probabilities of a categorical (or binary) dependent variable.
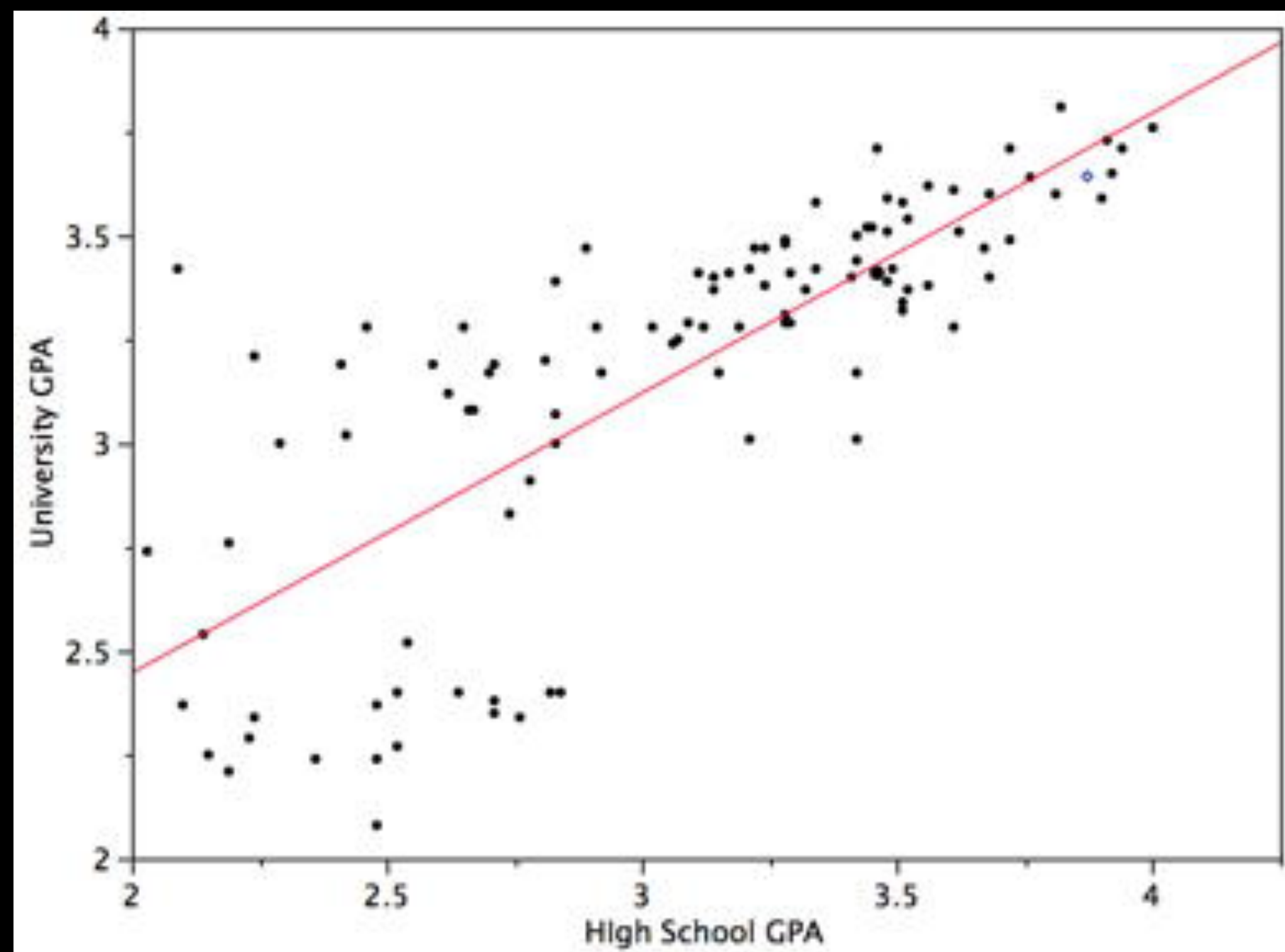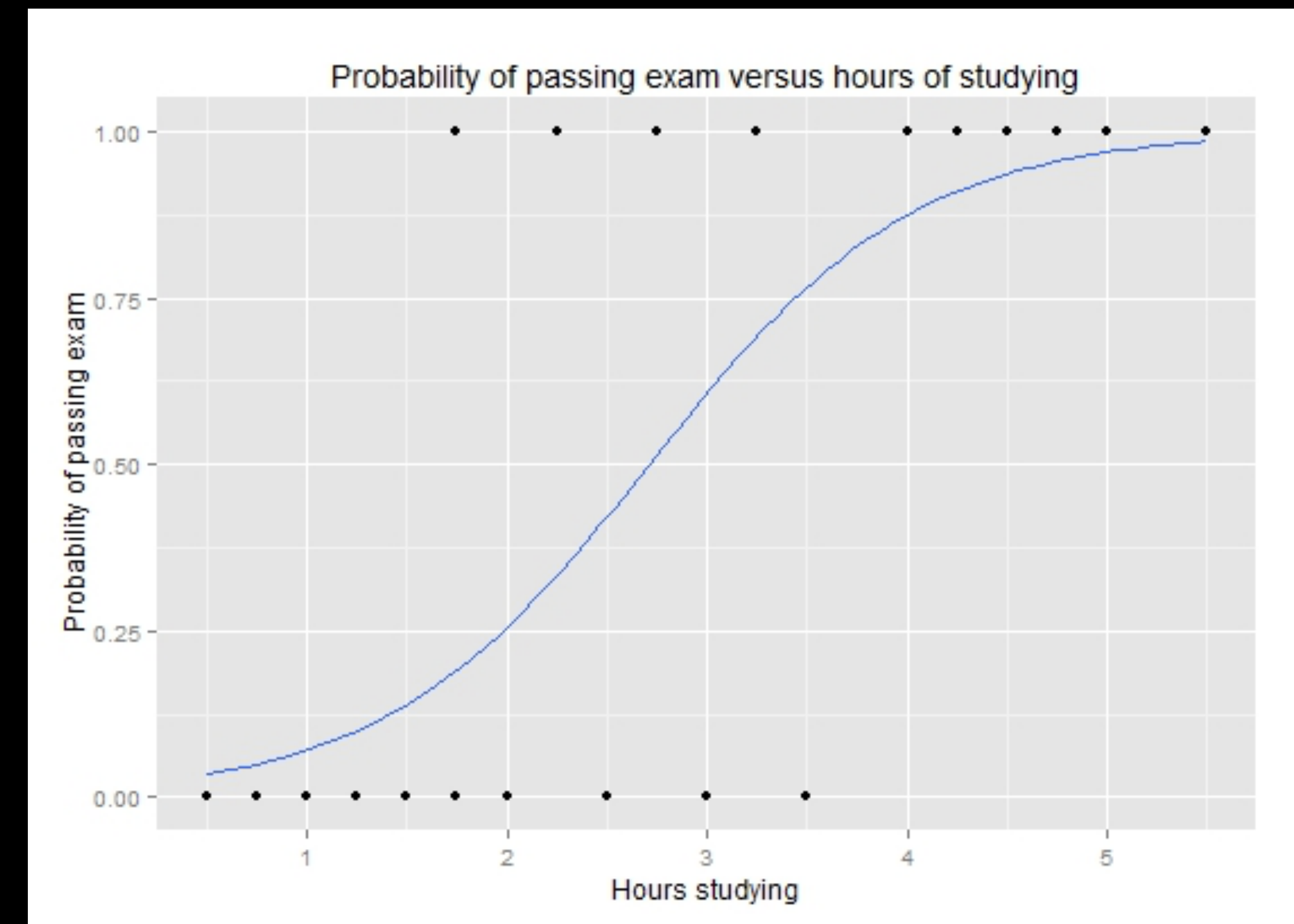
Two possible outcomes:
True/False
Success/Failure
Win/Lose
Adult/Child
...

LINEAR REGRESSION

LOGISTIC REGRESSION

Probability of passing exam versus hours of studying

LINEAR REGRESSION	
X	**y**
High School GPA	**University GPA**
3.5	4.0
3.3	3.8
2.5	**2.1**
4.0	4.0

LOGISTIC REGRESSION	
X	**y**
Hours Studied	**Passed/Failed**
5	**Pass**
5	**Fail**
15	Pass
25	Pass

Linear Regression	Logistic Regression
Predicts a continuous dependent variable.	Predicts a categorical dependent variable.
Finds line of "best fit" using **least squares**.	Finds the "sigmoid" curve using **maximum likelihood estimation**.
A linear relationship exists between the dependent and independent variables.	Linear relationship between the dependent and independent variables isn't mandatory.

MACHINE LEARNING 101: SOME CONCEPTUAL STUFF - I

- Training Data

- Test Data

- Cross Validation

- Overfitting

Logistic Regression: Toy Example

Logistic Regression: Demo I

Machine Learning 101:
Some Conceptual Stuff - II

- Confusion Matrix

- Precision

- Recall

WHAT IS A CONFUSION MATRIX?

		Actual Class	
		Condition Positive	Condition Negative
Predicted Class	**Predicted Positive**	True Positive	False Positive
	Predicted Negative	False Negative	True Negative

True Positive (TP):

- Reality: A wolf threatened.
- Shepherd said: "Wolf."
- Outcome: Shepherd is a hero.

False Positive (FP):

- Reality: No wolf threatened.
- Shepherd said: "Wolf."
- Outcome: Villagers are angry at shepherd for waking them up.

False Negative (FN):

- Reality: A wolf threatened.
- Shepherd said: "No wolf."
- Outcome: The wolf ate all the sheep.

True Negative (TN):

- Reality: No wolf threatened.
- Shepherd said: "No wolf."
- Outcome: Everyone is fine.

PRECISION

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

RECALL

$$Recall = \frac{TP}{TP + FN}$$