PSY9511: Seminar 2

The basics of regression and classification

Esten H. Leonardsen

26.02.2024

UNIVERSITETET
I OSLO

Outline

Today's lecture:

1. Recap of last lecture
2. Proposed solution for Assignment 1
3. Basics of regression and classification
    - Linear regression
    - K-nearest Neighbours
    - Logistic regression
4. Presentation of Assignment 2

Recap

What is statistical learning?

Recap

What is statistical learning?

- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

Recap

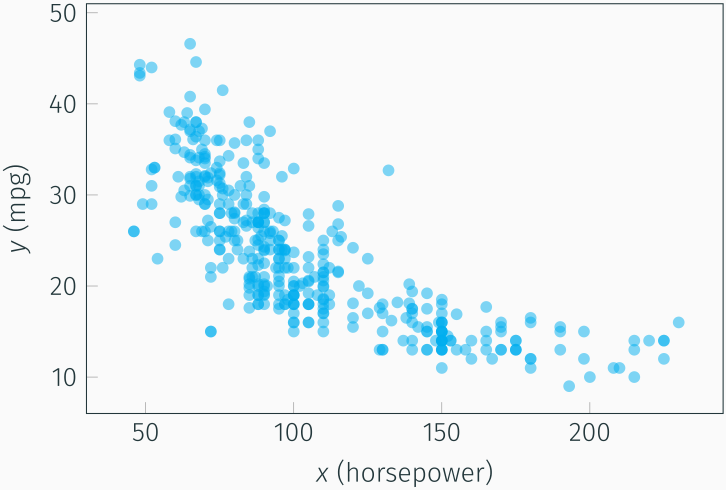
What is statistical learning?

- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.
- Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X, allows us to predict an output y.

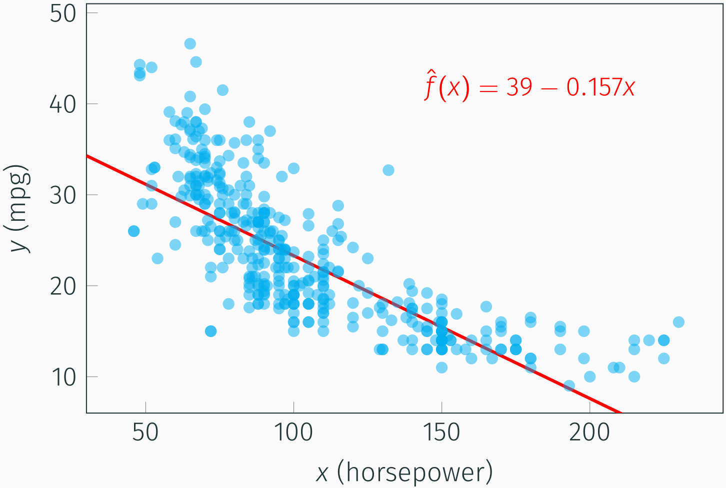
What is statistical learning?

- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.
- Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X, allows us to predict an output y.

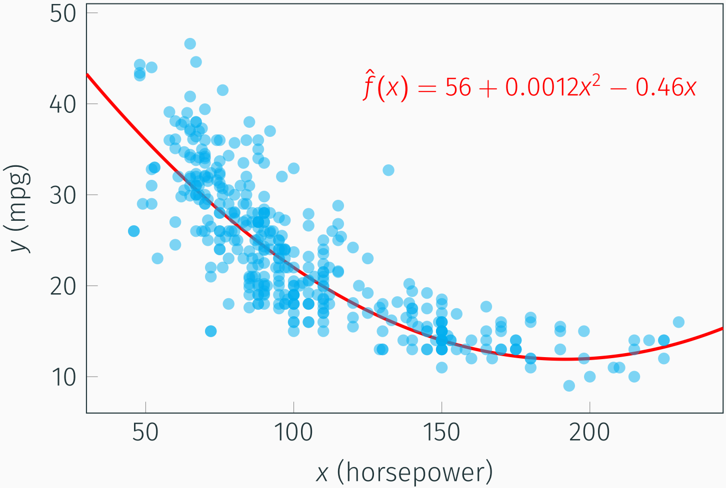
Recap



Recap

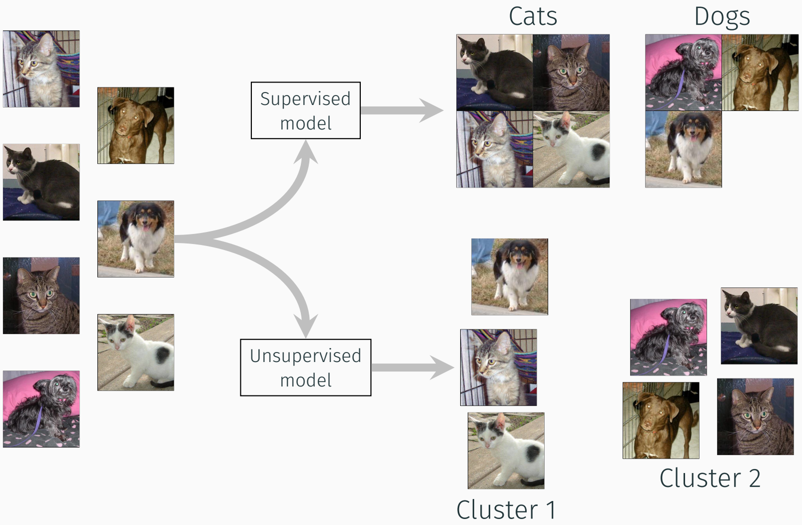


Recap



$\hat{f}(x) = 56 + 0.0012x^2 - 0.46x$
y (mpg)
x (horsepower)
50
40
30
20
10
50
100
150
200

Recap



Supervised model
Cats
Dogs
Unsupervised model
Cluster 1
Cluster 2

Recap

Regression

y
18
15
18
16
17

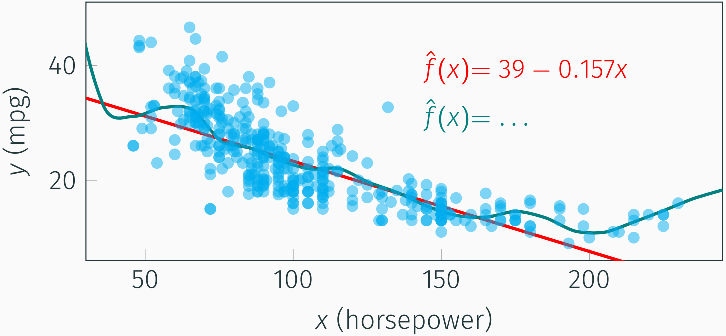
The predictive target y is a *continuous* (or *quantitative*) variable.

Classification

y
cat
cat
dog
cat
dog

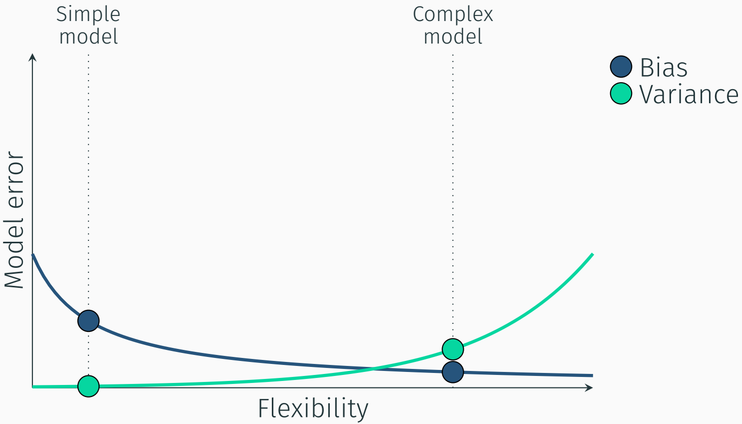
The predictive target y is a *categorical* (or *qualitative*) variable.

Recap



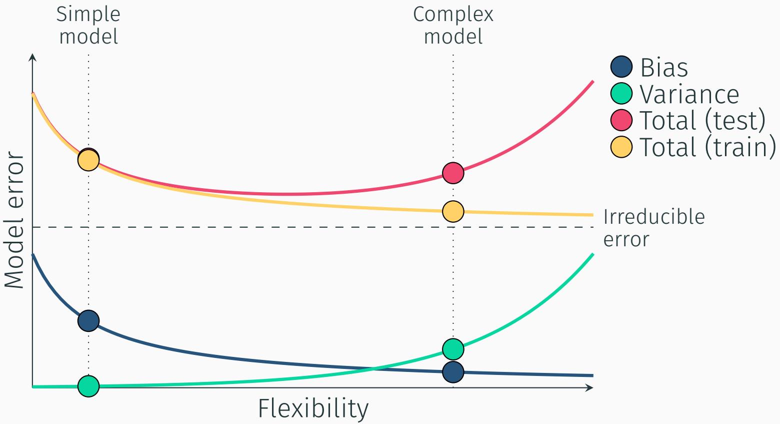
- **Parametric models** The function $\hat{f}(X)$ is relatively simple and can be described by a small number of parameters.
  - Linear regression: $\hat{f}(X) = \beta_0 + \beta_1 x$
- **Non-parametric models** The function $\hat{f}(X)$ is more complex and often relies directly on the data.

Recap



Simple model
Complex model
Bias
Variance
Model error
Flexibility

Recap



Simple model
Complex model
Bias
Variance
Total (test)
Total (train)
Irreducible error
Model error
Flexibility

The basics of regression and classification

UNIVERSITETET
I OSLO

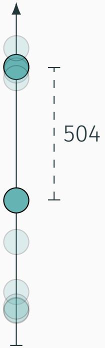
Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

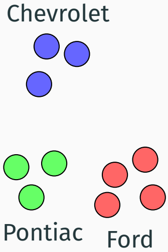
Regression vs. classification



Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet

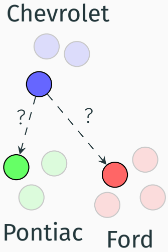
Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



Regression vs. classification

Mean squared error (MSE):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Accuracy:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(y_i, \hat{y}_i),$$

$$\mathbb{1}(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

Regression vs. classification

Regression:

- Predicting reaction time on a cognitive task based on sleep scores
- Predicting the age of an individual based on a brain scan
- Predicting anxiety scores based on questionnaire data

Classification:

- Predicting whether an individual is depressed based on cell phone usage data
- Predicting if a patient has dementia based on a brain scan
- Predicting whether a patient is happy based on their facial expression

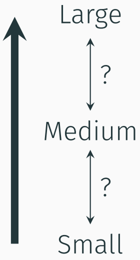
Regression vs. classification

Large

Medium

Small

Regression vs. classification

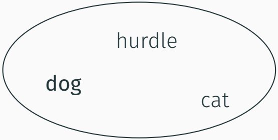


Regression vs. classification

The quick brown fox jumps over the lazy ______

Regression vs. classification

The quick brown fox jumps over the lazy ______



Classification

Regression vs. classification

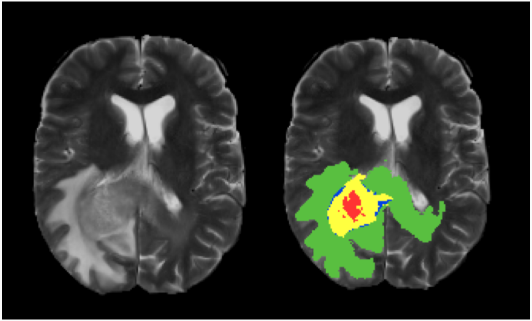
"Students taking a machine learning class" →

Regression vs. classification

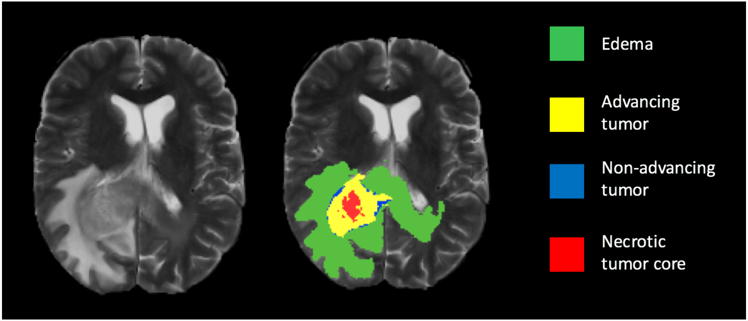
"Students taking a machine learning class" → → 

(128, 192, 64)

Regression vs. classification



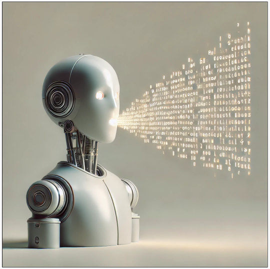
Regression vs. classification



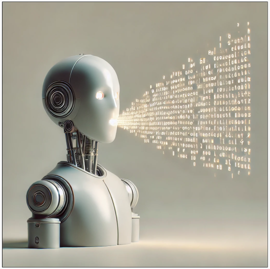
Regression vs. classification

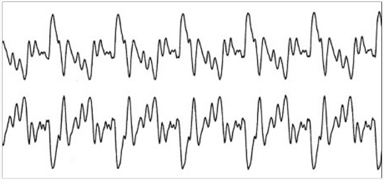
Regression vs. classification

Regression vs. classification



Regression vs. classification





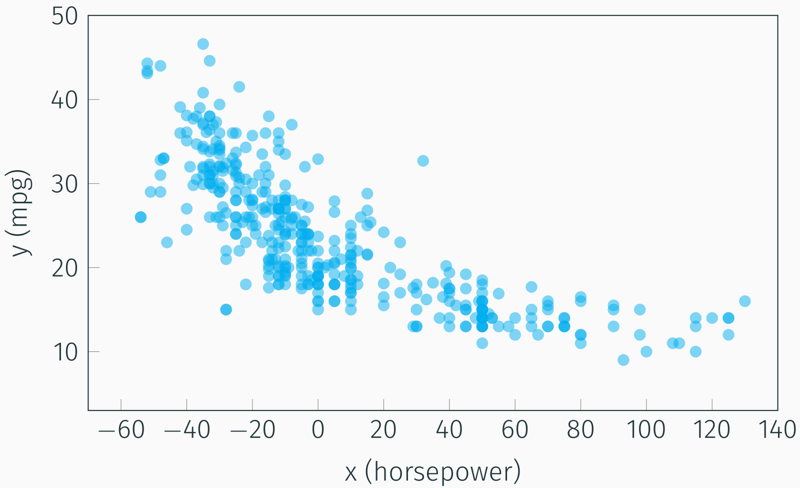
Regression vs. classification

Different types of outputs y require us to use different mathematical formulations of the problem we want to solve.

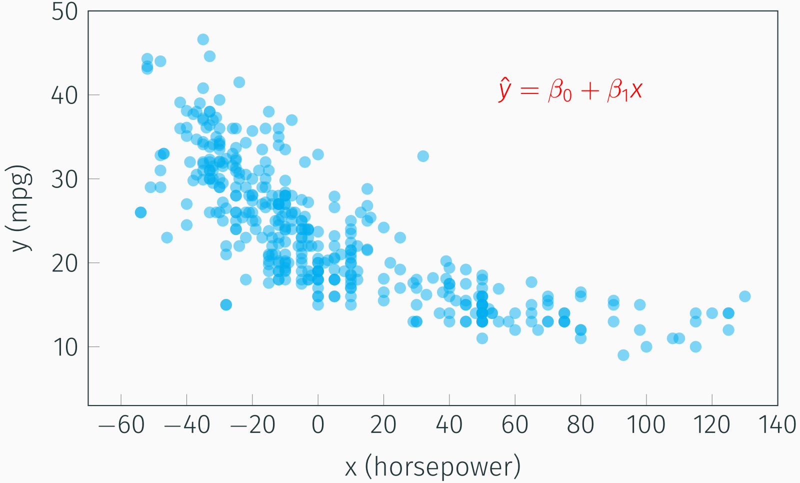
- Problems with quantitative outputs are solved via regression, often by minimizing the mean squared error
- Problems with qualitative outputs are solved by classification, often to maximize accuracy
- Ordinal regression falls between the two, with qualitative classes that have some kind of order
- A variety of other types of problems can be seen as special cases of these two

Linear regression (via ordinary least squares)

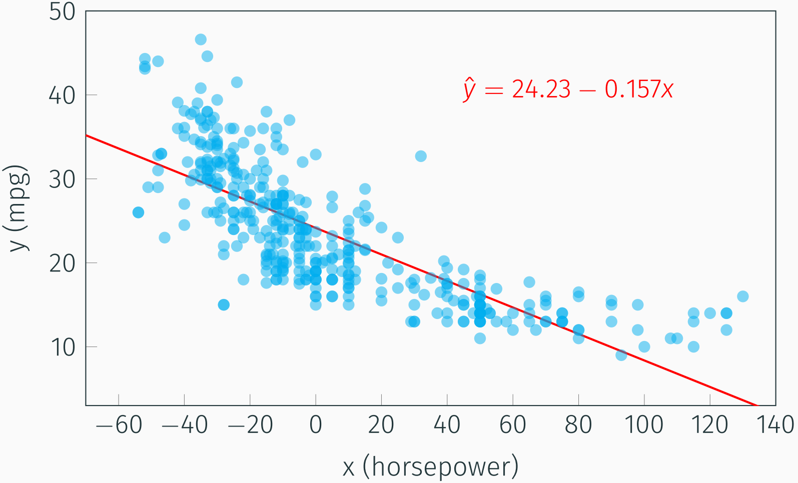
Linear regression (via ordinary least squares)



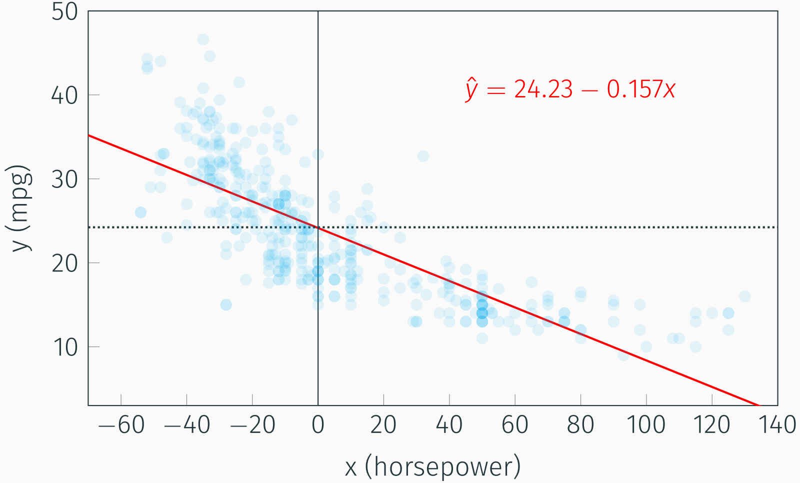
Linear regression (via ordinary least squares)



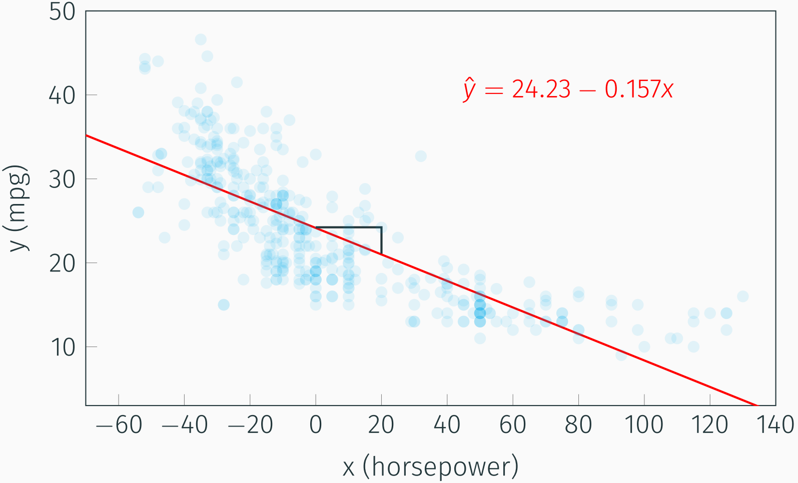
Linear regression (via ordinary least squares)



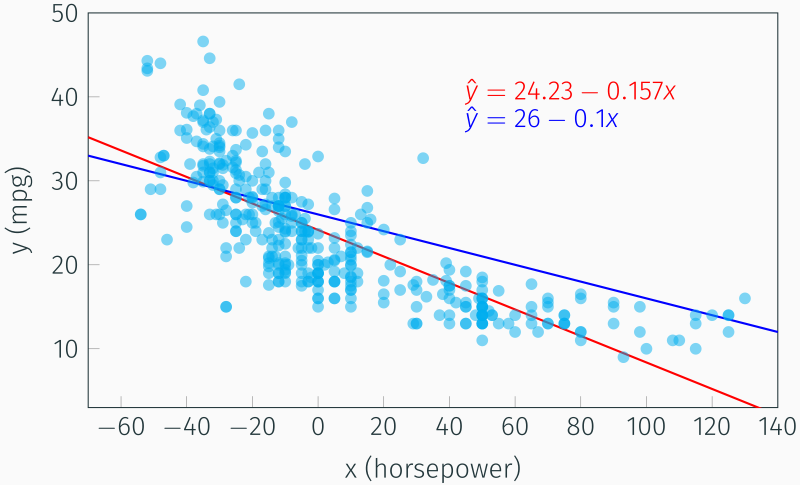
Linear regression (via ordinary least squares)



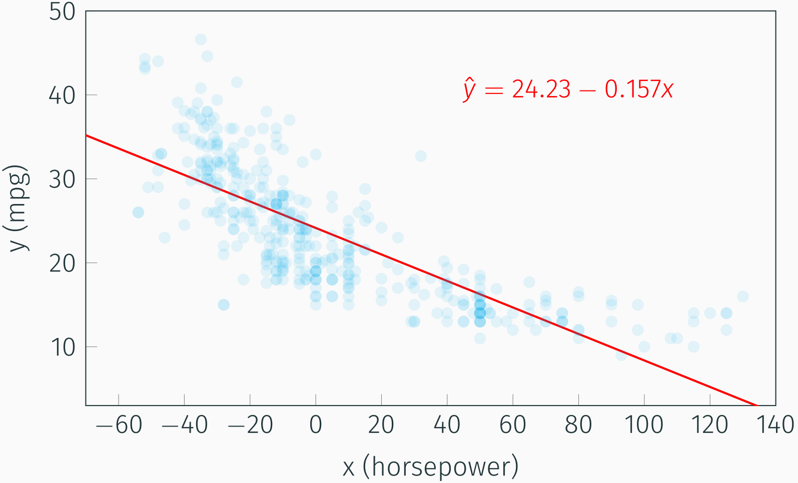
Linear regression (via ordinary least squares)



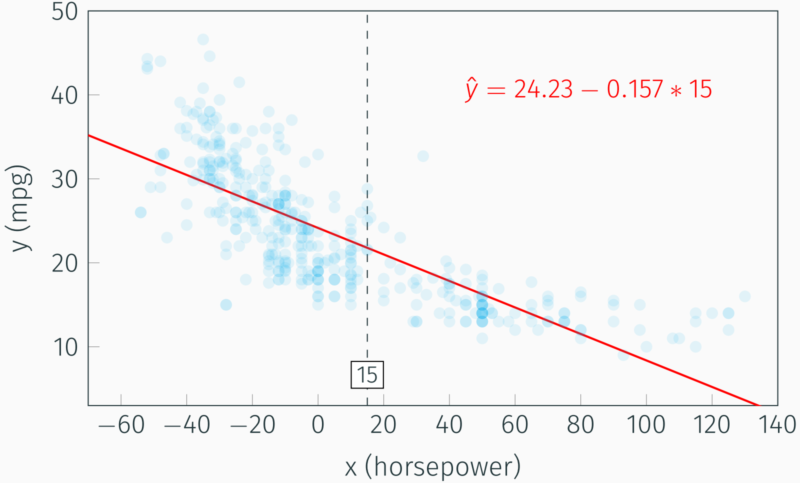
Linear regression (via ordinary least squares)



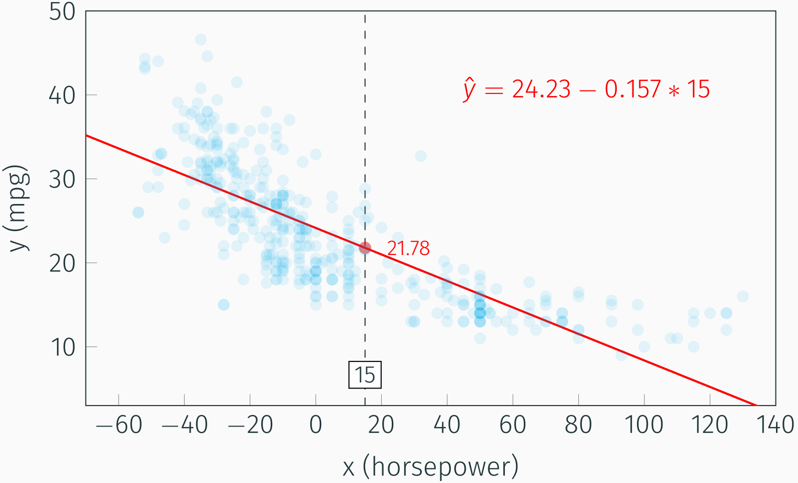
Linear regression (via ordinary least squares)



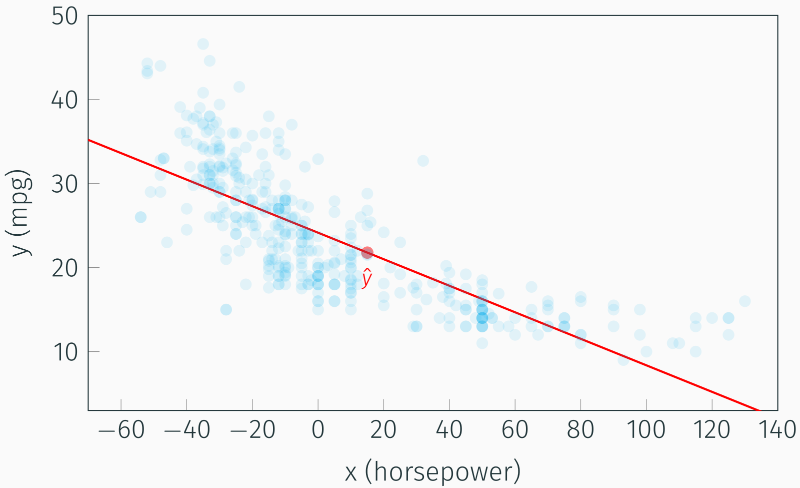
Linear regression (via ordinary least squares)



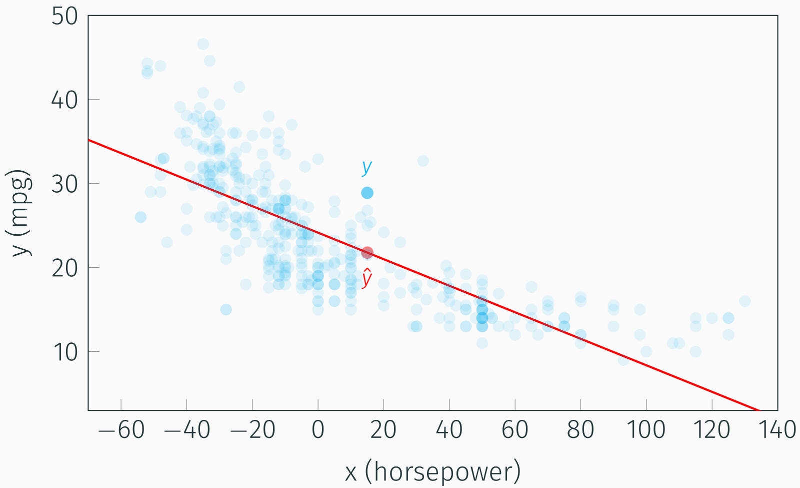
Linear regression (via ordinary least squares)



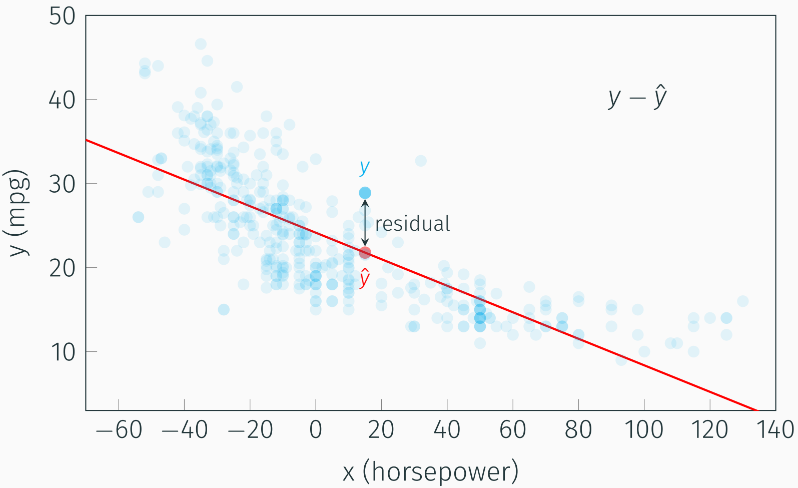
Linear regression (via ordinary least squares)



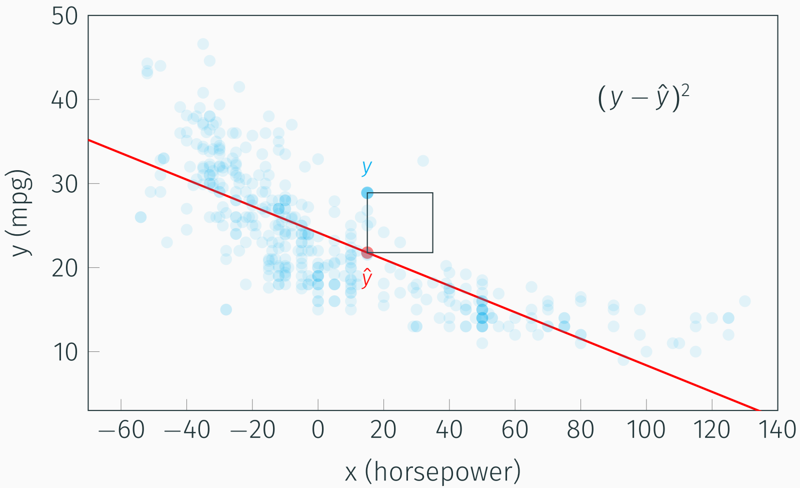
Linear regression (via ordinary least squares)



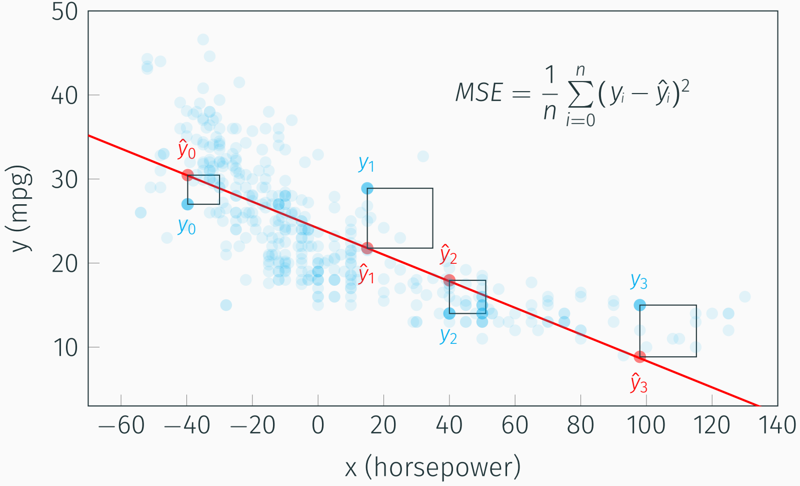
Linear regression (via ordinary least squares)



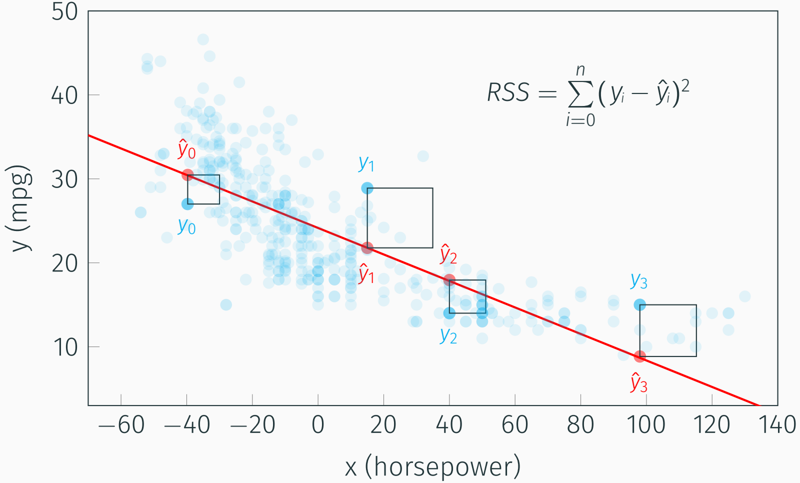
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

Linear regression: Models the relationship between input x and output y by finding the linear model $\hat{y} = \beta_0 + \beta_1 x$ that minimizes the residual sum of squares (RSS).

- β_0 refers to the intercept (or offset) of the model
- β_1 refers to the slope of the model

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 x$$

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 X$$

Fitting a linear regression model

$$\hat{y} = \beta_0 + \beta_1 x$$

Fitting a linear regression model

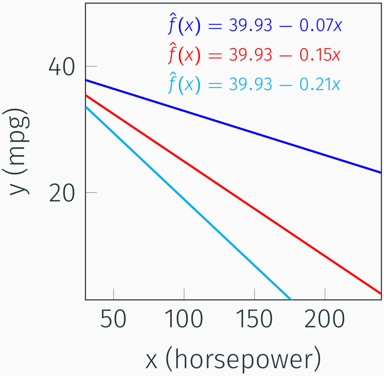
$$\hat{y} = \beta_0 + \beta_1 x$$

$$MSE = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2$$

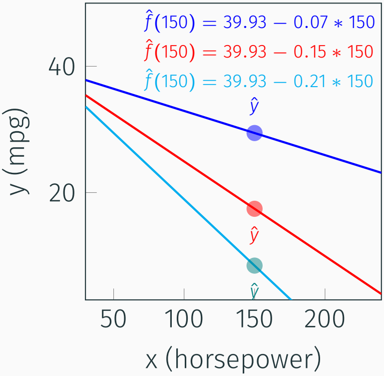
Fitting a linear regression model

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 + \beta_1 x_i)^2$$

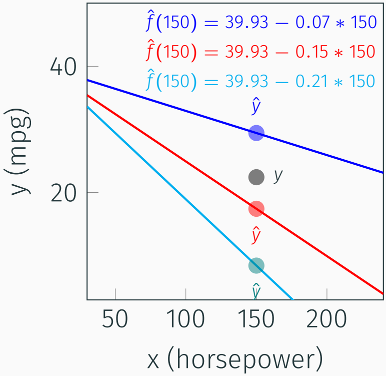
Fitting a linear regression model



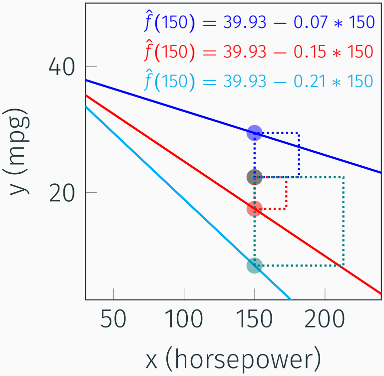
Fitting a linear regression model



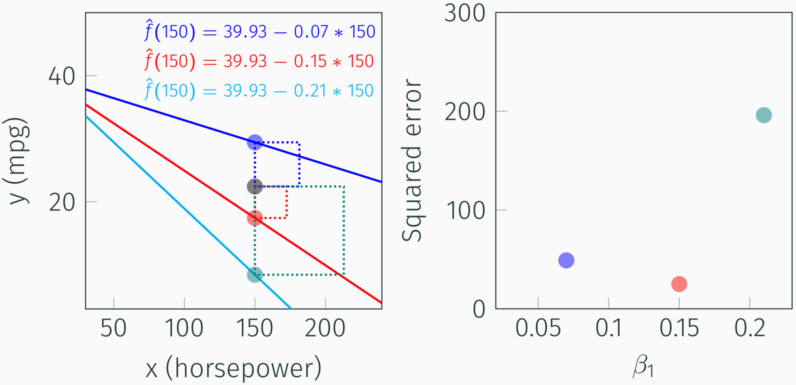
Fitting a linear regression model



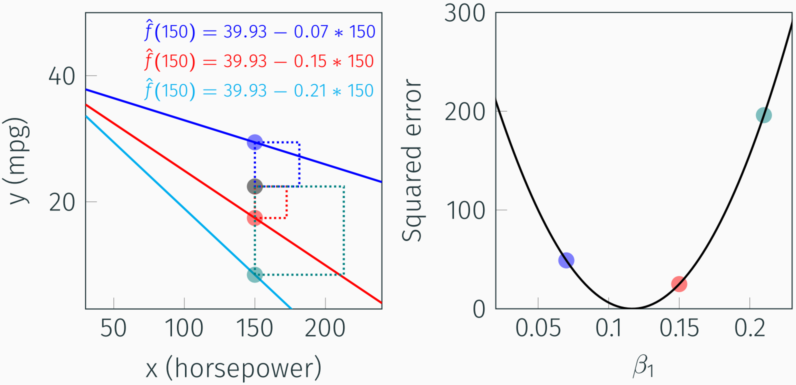
Fitting a linear regression model



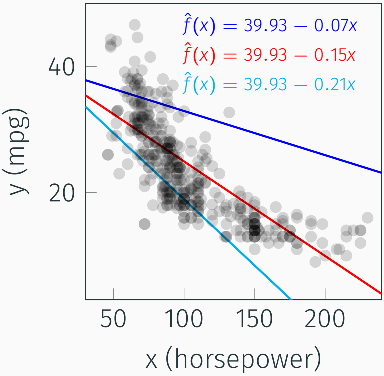
Fitting a linear regression model



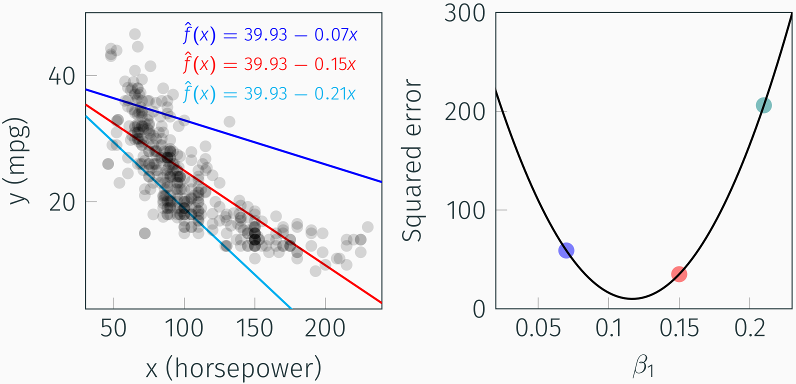
Fitting a linear regression model



Fitting a linear regression model



Fitting a linear regression model



Multivariate linear regression

$$\hat{f}(x) = \beta_0 + \beta_1 x_1$$

Multivariate linear regression

$$\hat{f}(x) = \beta_0 + \beta_1 x_1$$
$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

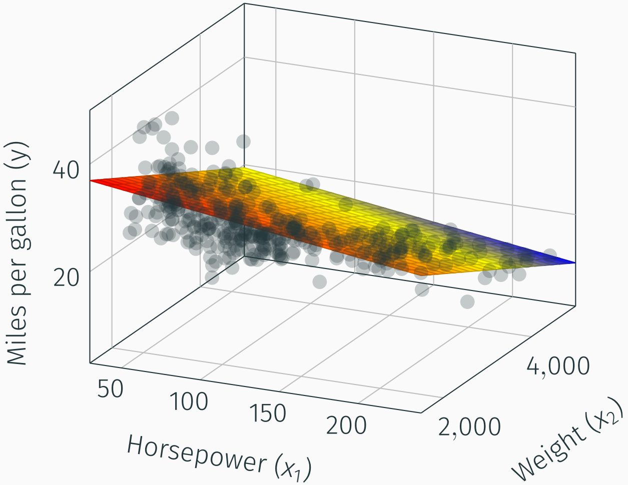
Multivariate linear regression

$$\hat{f}(x) = \beta_0 + \beta_1 x_1$$

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

$$\hat{f}(X) = \beta_0 + \sum_{i=0}^{p} \beta_i X_i$$

Multivariate linear regression



Categorical variables

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Chevrolet
17	Ford
15	Ford
32	Chevrolet
14	Ford
14	Ford
28	Chevrolet

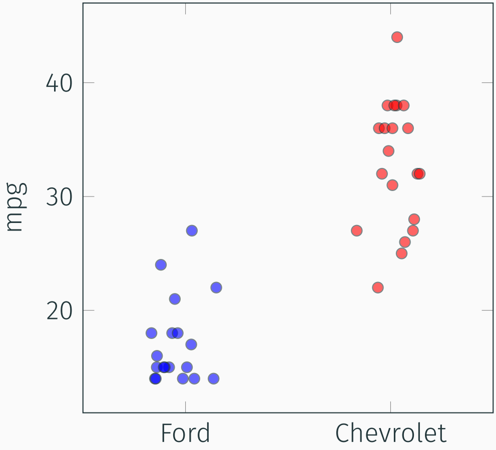
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$

Categorical variables

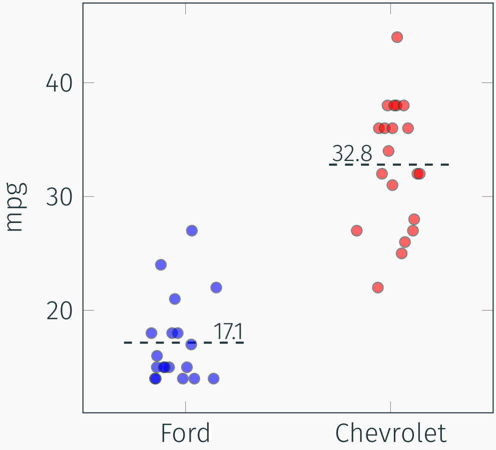
mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet}$$

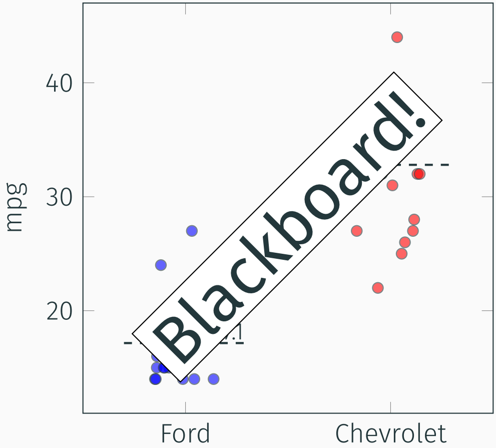
Categorical variables



Categorical variables



Categorical variables



Categorical variables

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Pontiac
17	Ford
15	Ford
32	Pontiac
14	Ford
14	Pontiac
28	Chevrolet

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$

Categorical variables

mpg	manufacturer	chevrolet	pontiac
36	Chevrolet	1	0
15	Ford	0	0
25	Chevrolet	1	0
26	Pontiac	0	1
17	Ford	0	0
15	Ford	0	0
32	Pontiac	0	1
14	Ford	0	0
14	Pontiac	0	1
28	Chevrolet	1	0

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{pontiac}$$

Categorical variables

In[1]:

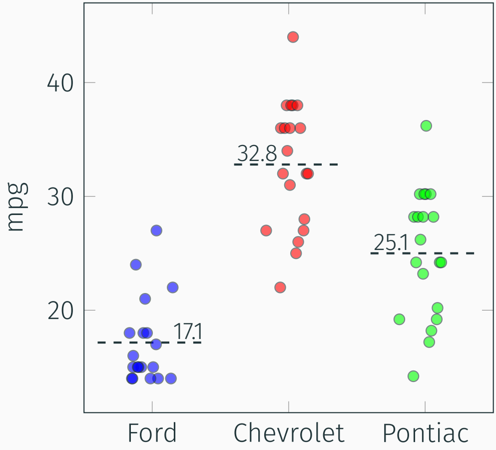
```
import pandas as pd

df = pd.DataFrame(...)
print(f'Columns before: {df.columns.values}')
df = pd.get_dummies(df)
print(f'Columns after: {df.columns.values})
```

Out[1]:

```
Columns before: ['manufacturer']
Columns after: ['manufacturer_chevrolet' 'manufacturer_ford']
```

Categorical variables



Categorical variables

mpg	chevrolet	horsepower
36	1	130
15	0	165
25	1	150
26	1	150
17	0	140
15	0	198
32	1	220
14	0	215
14	0	225
28	1	212

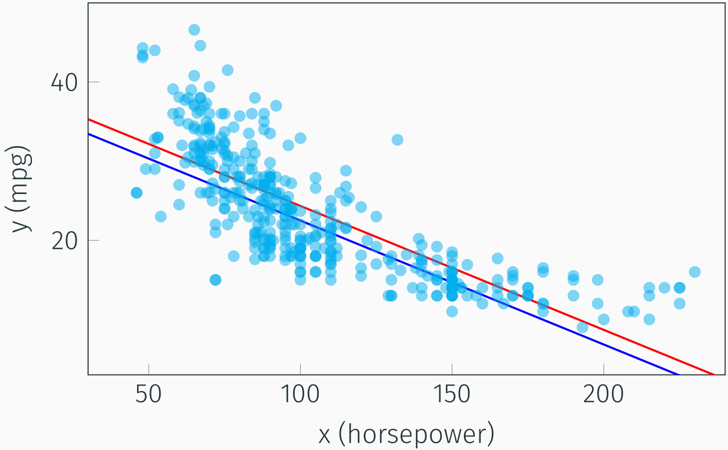
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{horsepower}$$

Categorical variables

mpg	chevrolet	horsepower
36	1	130
15	0	165
25	1	150
26	1	150
17	0	140
15	0	198
32	1	220
14	0	215
14	0	225
28	1	212

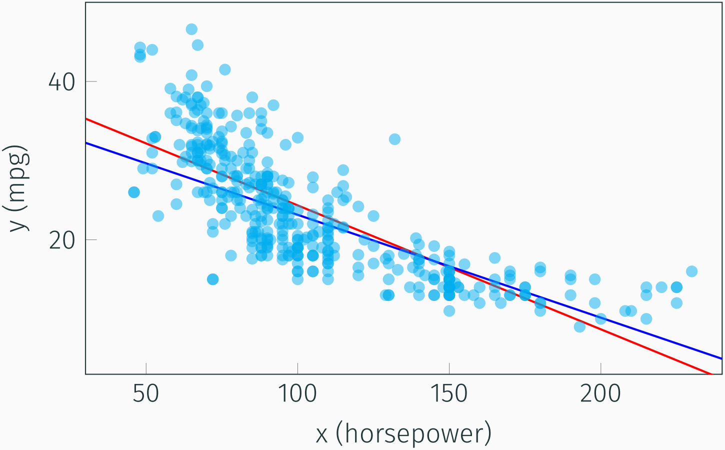
$$\widehat{mpg} = \begin{cases} \beta_0 + \beta_1 + \beta_2 \times \text{horsepower} & \text{if chevrolet} \\ \beta_0 + \beta_2 \times \text{horsepower} & \text{else} \end{cases}$$

Categorical variables

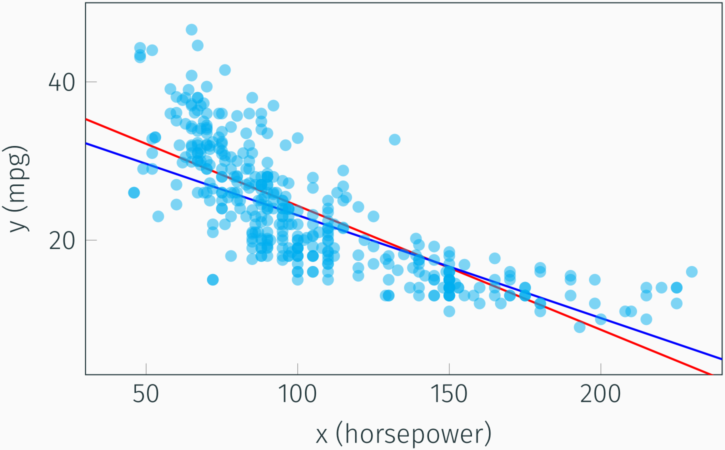


$$\widehat{mpg} = \begin{cases} \color{blue}{\beta_0 + \beta_1 + \beta_2 \times \text{horsepower}} & \text{if chevrolet} \\ \color{red}{\beta_0 + \beta_2 \times \text{horsepower}} & \text{else} \end{cases}$$

Categorical variables

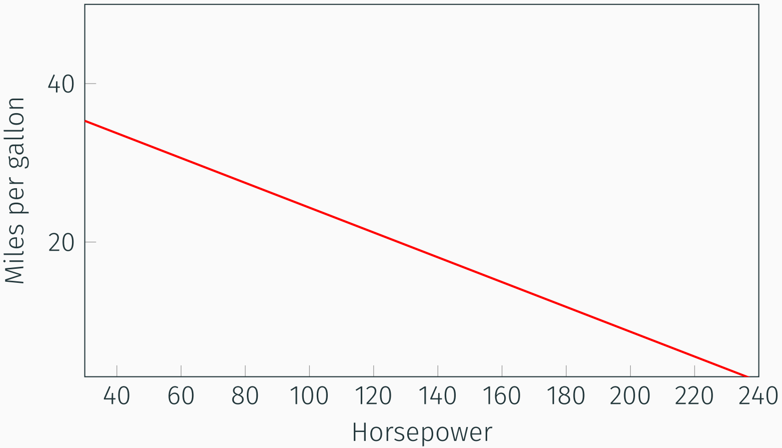


Categorical variables

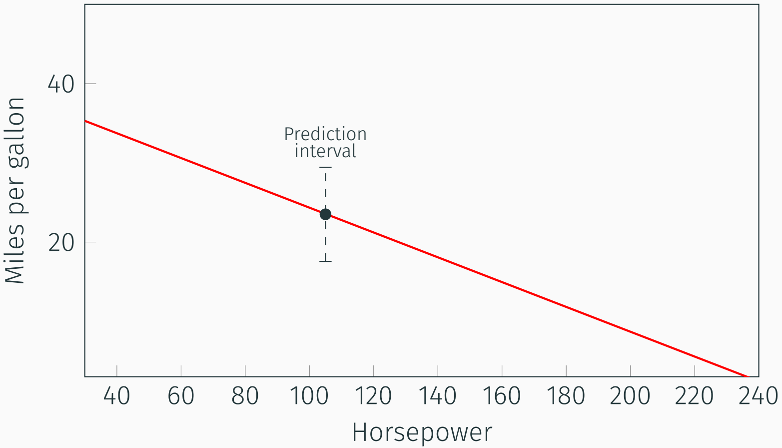


$$\widehat{mpg} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{horsepower}$$
$$+\beta_3 \times \text{chevrolet} \times \text{horsepower}$$

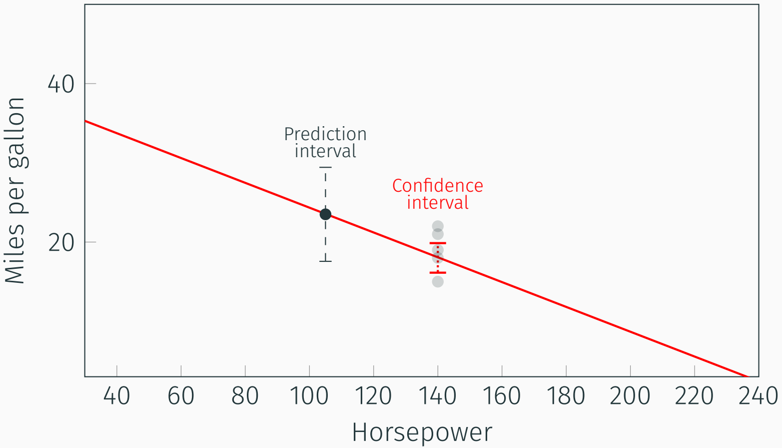
Confidence intervals



Confidence intervals



Confidence intervals



Confidence intervals

```
predict(fit, newdata=data.frame(horsepower=105),
        interval='prediction', level=0.95)
```

```
     fit    lwr    upr
1 23.535 17.158 29.912
```

```
predict(fit, newdata=data.frame(horsepower=105),
        interval='confidence', level=0.95)
```

```
     fit    lwr    upr
1 23.535 23.023 24.047
```

Confidence intervals

In[1]:

```
import statsmodels.api as sm

model = sm.OLS(df['mpg'], sm.add_constant(df[['horsepower']]))
fit = model.fit()
new_input = sm.add_constant(pd.DataFrame({'horsepower': [105, 106]}))
intervals = fit.get_prediction(new_input).summary_frame()
print(intervals)
```

Out[1]:

```
mean mean_se mean_ci_lower mean_ci_upper obs_ci_lower obs_ci_upper
0 24.467077 0.251262 23.973079 24.961075 14.809396 34.124758
1 31.096556 0.398740 30.312607 31.880505 21.419710 40.773402
```

Confidence intervals

Why do we need both of these?
Where are they useful (e.g. what is most useful in a scientific publication versus a business setting)?

Confidence intervals

http://localhost:8888/notebooks/notebooks%2FLinear%20regression.ipynb

Linear regression: Summary

Linear regression: The workhorse of machine learning

- Models the relationship between (either singular or mulitple) inputs X and (a continuous) output y as a linear function
  - Inputs can be both continuous and categorical
- A strict parametric form limits the expressivity of the model
  - More advanced terms can be explicitly added
  - The strictness allows for extended functionality, such as computing confidence intervals
  - **Makes the model human interpretable**

K-Nearest Neighbours

K-Nearest Neighbours

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

K-Nearest Neighbours

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

K-Nearest Neighbours

Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

K-Nearest Neighbours

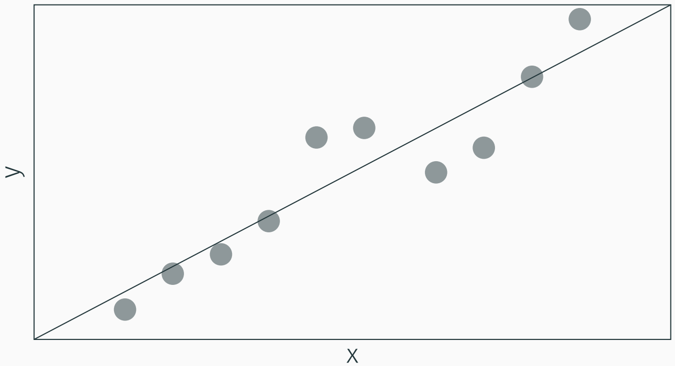
Linear regression:

$$\hat{f}(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

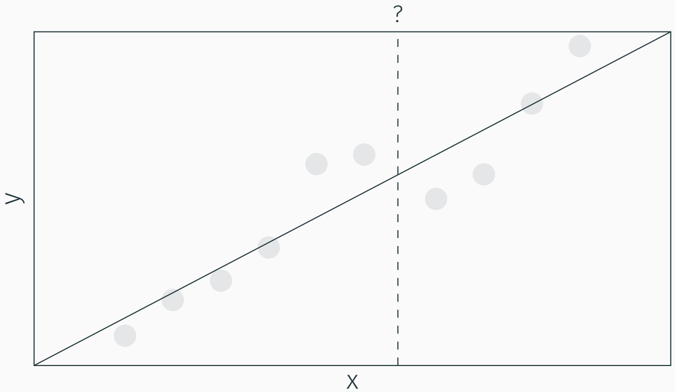
K-Nearest Neighbours:

$$\hat{f}(X) = \frac{1}{K} \sum_{x_i \in \mathcal{N}} y_i$$

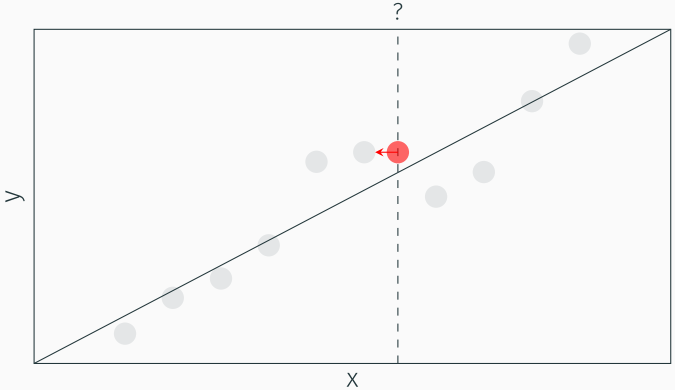
K-Nearest Neighbours



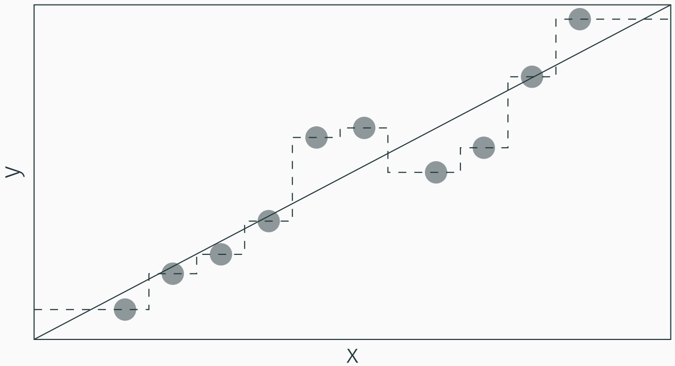
K-Nearest Neighbours



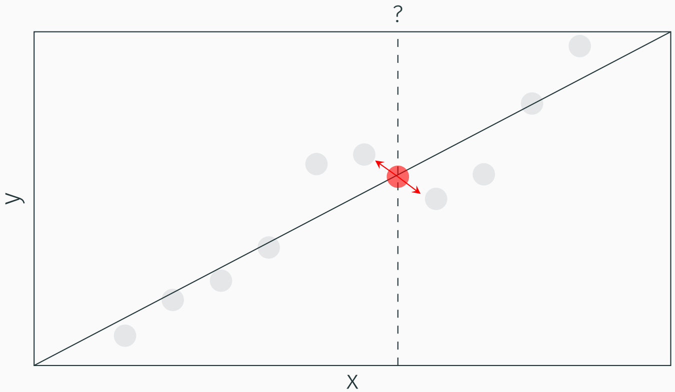
K-Nearest Neighbours



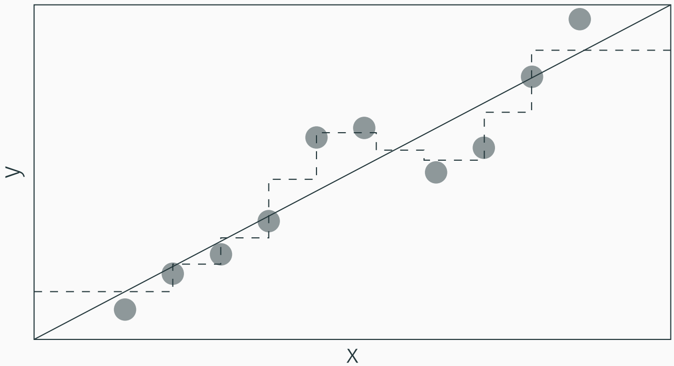
K-Nearest Neighbours



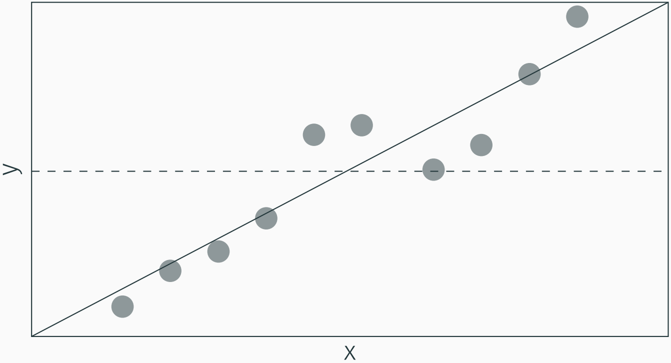
K-Nearest Neighbours



K-Nearest Neighbours



K-Nearest Neighbours



K-Nearest Neighbours

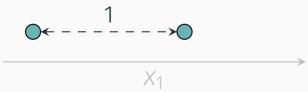
How does the bias-variance trade-off relate to the choice of K?

K-Nearest Neighbours

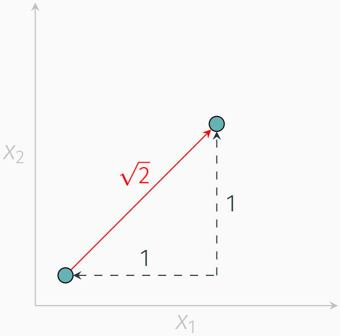
The curse of dimensionality:

Things become complicated in very high dimensions

K-Nearest Neighbours



K-Nearest Neighbours



K-Nearest Neighbours

K-Nearest Neighbours: An intuitive model relying on similarities between datapoints to make predictions

- No assumptions about the functional relationship between inputs and outputs
- Directly trades off bias and variance through the choice of K
- Does not work well in high dimensions as space gets more sparsely populated, yielding less dense neighbourhoods
- Can be fidgety in practice (e.g. how should one choose the scales of different predictors), not very commonly used

Logistic regression

Logistic regression

mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

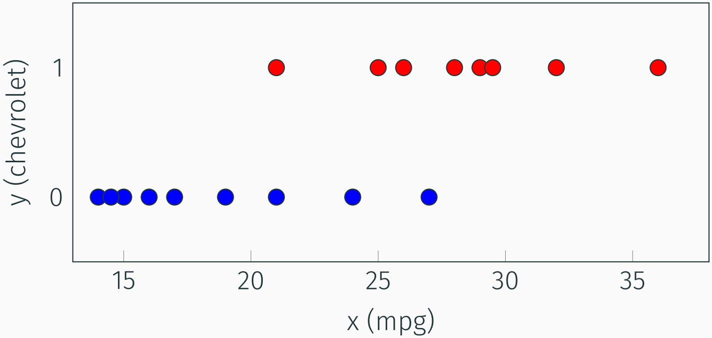
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet}$$

Logistic regression

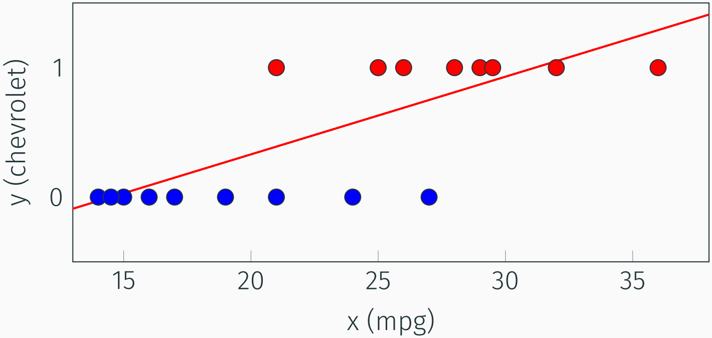
mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

$$\widehat{\text{chevrolet}} = \beta_0 + \beta_1 \times \text{mpg}$$

Logistic regression

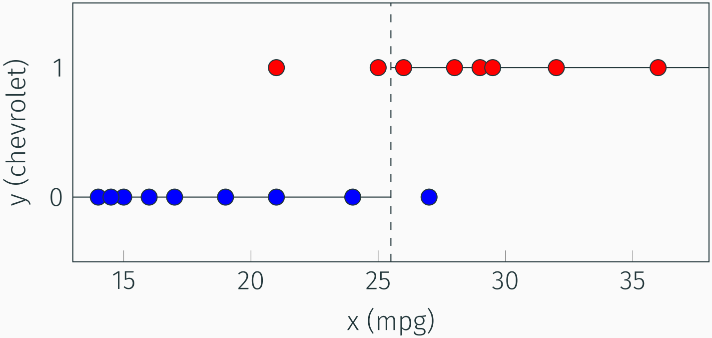


Logistic regression

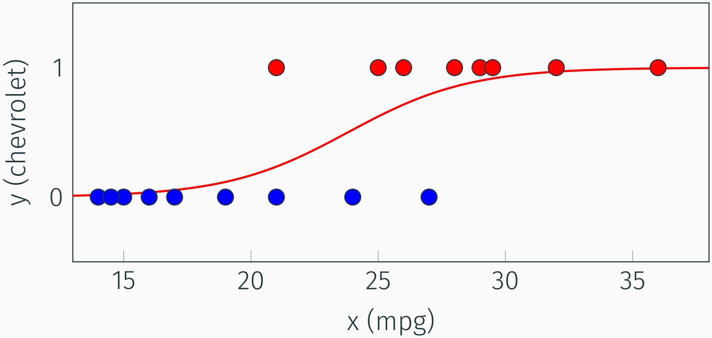


$$\widehat{\text{chevrolet}} = -0.87 + 0.06 \times \text{mpg}$$

Logistic regression



Logistic regression



$$\widehat{\text{chevrolet}} = \frac{e^{-10.22+0.42\times \text{mpg}}}{1+e^{-10.22+0.42\times \text{mpg}}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$e^{\beta_0+\beta_1 x} \rightarrow 0$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$e^{\beta_0+\beta_1 x} \to 0 \implies \hat{y} = 0$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$\begin{aligned} e^{\beta_0+\beta_1 x} &\to 0 \quad \implies \hat{y} = 0 \\ e^{\beta_0+\beta_1 x} &\to \infty \end{aligned}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$\begin{aligned} e^{\beta_0+\beta_1 x} \to 0 &\implies \hat{y} = 0 \\ e^{\beta_0+\beta_1 x} \to \infty &\implies \hat{y} = 1 \end{aligned}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$\begin{aligned} e^{\beta_0+\beta_1 x} \to 0 \quad &\implies \hat{y} = 0 \\ e^{\beta_0+\beta_1 x} \to \infty \quad &\implies \hat{y} = 1 \end{aligned}$$

$$0 \leq \hat{y} \leq 1$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$\begin{aligned} e^{\beta_0+\beta_1 x} \to 0 &\implies \hat{y} = 0 \\ e^{\beta_0+\beta_1 x} \to \infty &\implies \hat{y} = 1 \end{aligned}$$

$$0 \leq \hat{y} \leq 1 \implies \hat{y} = Pr(Y = 1|X)$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 x$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 x$$

"... counterintuitive and challenging to interpret." - James Jaccard

Logistic regression

$$\hat{y} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

More than one class?

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

More than one class?

$$Pr(Y = k|X = x) = \frac{e^{\beta_{0k}+\beta_{1k} x}}{\sum_{l=1}^{K} e^{\beta_{0l}+\beta_{1l} x}}$$

Logistic regression

$$\hat{y} = \frac{e^{\beta_0+\beta \dots}}{1+ \dots}$$

More th… class?

Blackboard!

$$Pr(Y \dots k|X = x) = \frac{e^{\beta_{0k}+\beta_{1k}x}}{\sum_{l=1}^{K} e^{\beta_{0l}+\beta_{1l}x}}$$

Logistic regression

Logistic regression: Extends linear regression to classification.

- Treats all members of a class (approximately) equally.
- Ouputs an understandable quantity: The probability of a sample belonging to the positive class.
- Somewhat interpretable, although not as much as linear regression
- Can trivially be extended to include multiple classes

Logistic regression

http://localhost:8888/notebooks/notebooks%2FLogistic%20regression.ipynb

Classification metrics

UNIVERSITETET
I OSLO

Classification metrics

The two most common, severe, mistakes made in machine learning studies in psychology and neuroscience (in my opinion) are:

- Poor validation and testing strategies (Lecture 4)
- Using incorrect performance measures, most commonly accuracy (Blackboard!)

Assignment 2

1. Download the Auto.csv-dataset
2. Fit a single linear regression model
3. Fit a multivariate linear regression model
4. Fit a logistic regression model