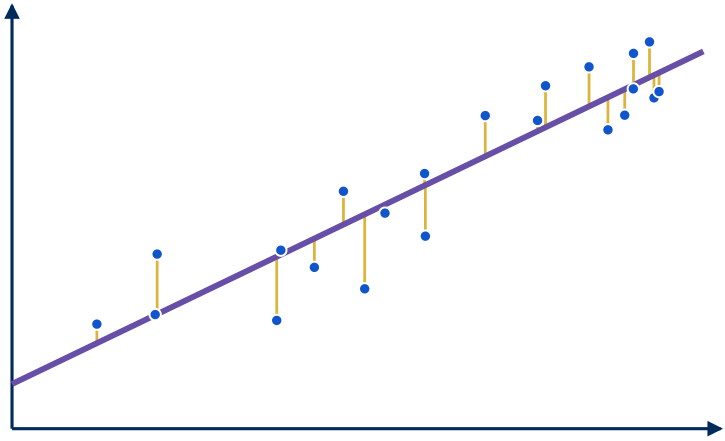
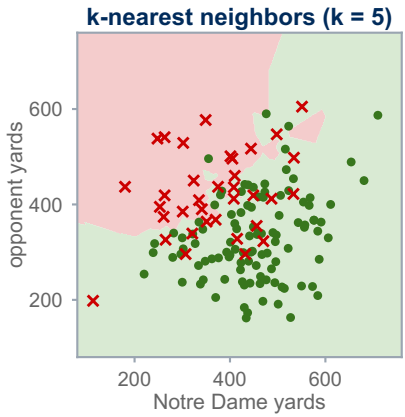


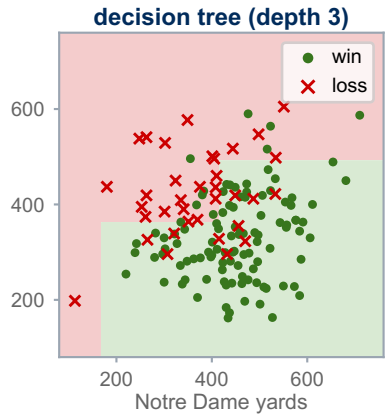
UNIT 02 · CHAPTER 5

Linear Regression

CSE 30124 · Introduction to Artificial Intelligence
William Theisen · University of Notre Dame

Classification Tasks





Regression Tasks

Quick check (old exam question): is a regression task supervised or unsupervised learning?

Where "Regression" Comes From

In 1886 Francis Galton compared the heights of 928 adult children with the heights of their parents (205 families).

Tall parents had tall children, but the children were **usually closer to average than their parents**, and the same was true for short parents. Galton called this **"regression towards mediocrity"**.

The ___ he drew through the data kept the name, which is why fitting a line is called **linear regression** even when nothing is regressing.

The Simplest Model: A Line

Linear Regression

Unfortunately, because linear regression is so old, **it has a lot of vocab specific to it** that has accrued and changed over time, depending on the field using it. This can make it tricky to recognize a linear regression in different contexts.

[5.1] a line with one feature

$$\hat{y} = w_1 x_1 + w_0$$

- $\hat{y}$: _____
- w_1, w_0: _____
- x_1: _____

Finding Our Line

How do we actually find the line that "fits" the data the best?

Cost and Loss Functions

sum of squared errors

$$[5.2] \quad SSE(w) = \sum_i (y_i - \hat{y}_i)^2$$

mean squared error

$$[5.3] \quad MSE(w) = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2$$

Why Squares, Again

Maximum likelihood estimation (MLE) is a method of estimating the parameters of an _____, given some observed data.

Optimization Problems

Visualizing Loss Functions with Respect to the Weights

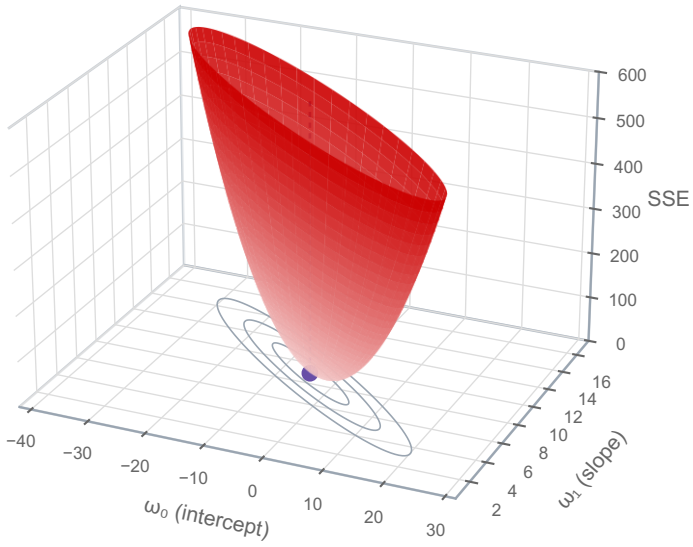
We can plot the loss function with respect to ω_1:

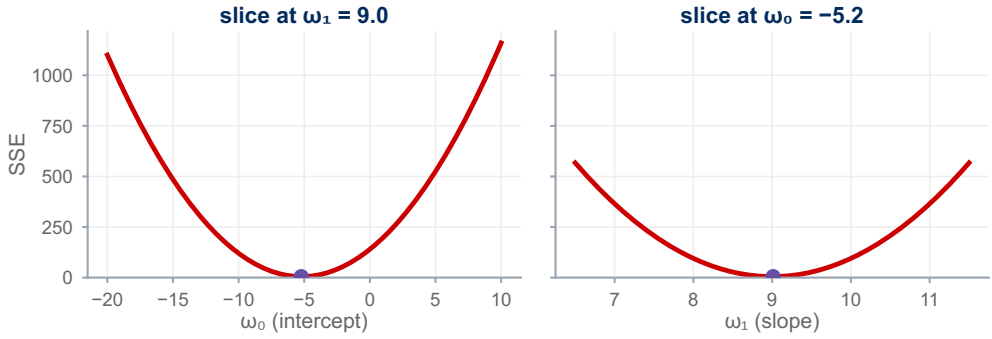


Finding the Bottom

So given a loss function, our goal is to _____ it. We do this by finding the value of ω_1 at the bottom of the bowl.

Two Weights at Once





That's two equations to solve together, and with ten features it would be eleven. Writing each one out by hand gets old fast.

Matrix Notation for Linear Regressions

We can instead represent our entire problem as matrices and matrix operations.

Ordinary Least Squares

Ordinary Least Squares, Continued

Worked OLS Example

game	yards	x_1	y (points)
2019 Michigan	180	2	14
2012 Michigan State	300	3	20
2013 Purdue	400	4	31
2019 Boston College	501	5	40
2017 Temple	606	6	49

Give X a **column of 1s** for the intercept, then follow the OLS recipe.

So $\hat{y} = 9.0\,x_1 - 5.2$: about _____.

What the Line Learned

Quick check: in `2018` Notre Dame gained `550` yards against Stanford. What does this line predict, and what is the residual if Notre Dame scored `38`?

OLS in sklearn

```python
import pandas as pd
from sklearn.linear_model import LinearRegression

games = pd.read_csv("nd_football.csv")
X = games[["nd_total_yards"]] / 100   # hundreds
y = games["nd_points"]

model = LinearRegression().fit(X, y)
print(model.coef_.round(2), model.intercept_.round(2))
print(round(model.score(X, y), 2))    # R²
```

OUTPUT

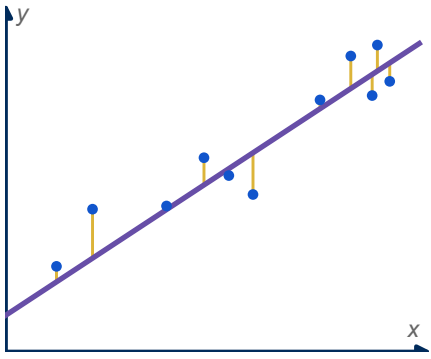
```
[9.03] -6.73
0.55
```

Multiple Linear Regression

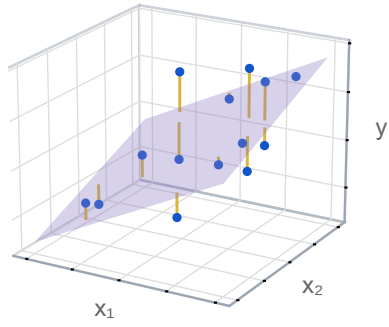
When we perform a linear regression on a dataset with more than one feature, we call this a _____.

Hyperplanes

1 feature: a line



2 features: a plane



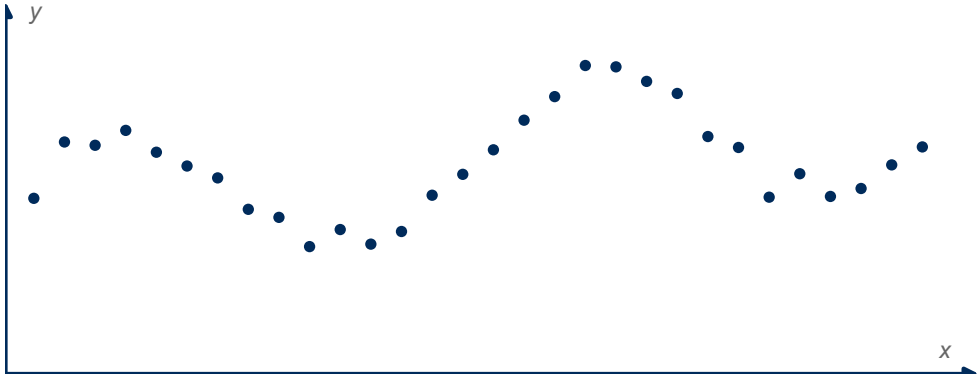
With one feature the model is a line, with two it's a _____ and with more than two it's a _____.

Quick check (old exam question): what would the equation look like for a multiple linear regression with three features?

Multiple Linear Regressions

Representing Curves

Not every dataset is best represented with a straight line. How could you use a **linear regression** to model a sine wave?

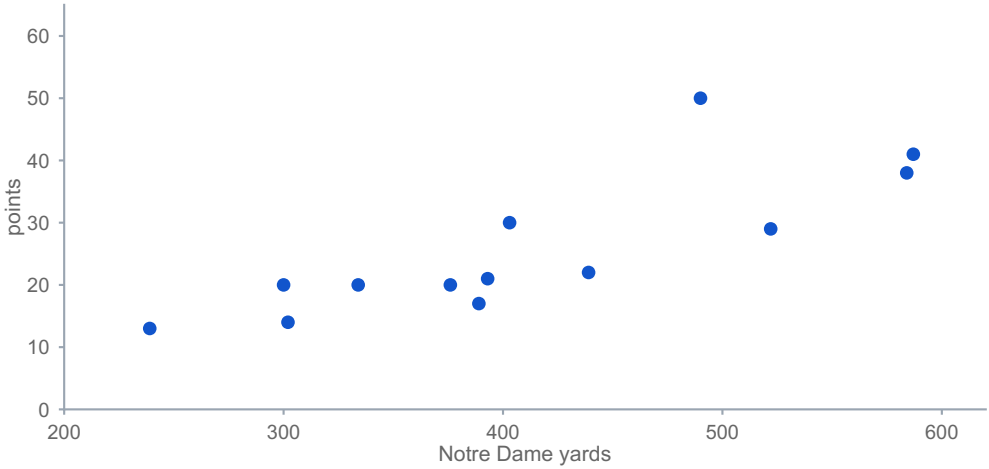


A linear regression is only linear in its _____, not in its inputs.

Polynomial Regression

Overfitting

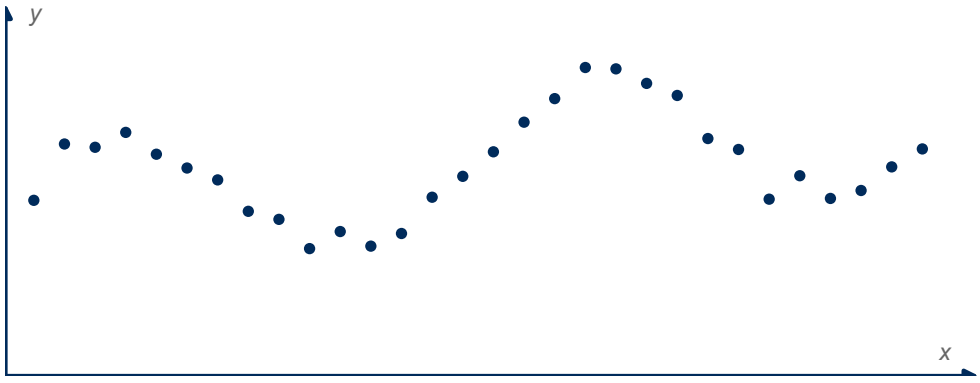
Overfitting



Variance

Underfitting

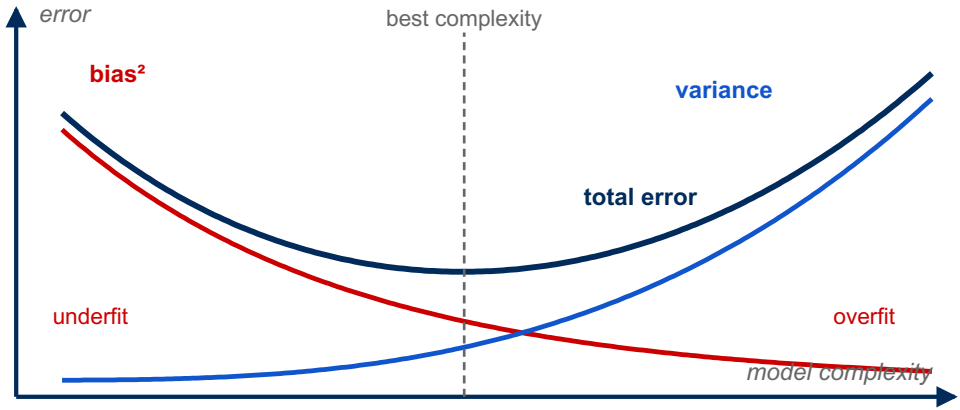
Bias



degree 8: nine weights

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 + w_6 x^6 + w_7 x^7 + w_8 x^8$$

The Bias-Variance Tradeoff



The **bias–variance tradeoff** describes the relationship between a model's complexity, the accuracy of its predictions (measured as overall error), and how well it can make predictions on unseen data.

Issues with Inverses