General Instruction Review

AI4ALL NLP Group

Comparisons

- Supervised vs Unsupervised Learning
- Regression vs Classification
- Underfitting vs Overfitting

Linear and Logistic Regression

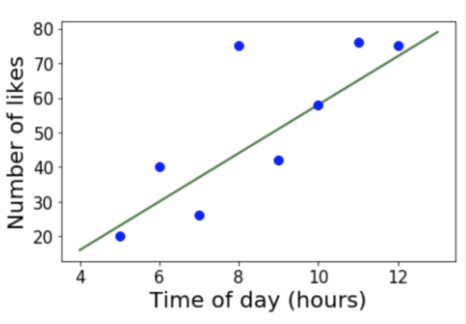
Goal: Predict numerical outputs from previously unseen inputs.

Linear and Logistic Regression

Goal: Predict numerical outputs from previously unseen inputs.

Linear regression:

- Models: Linear relationship between inputs and outputs.
- Find the line ($\hat{y} = mx + b$) that minimizes error.



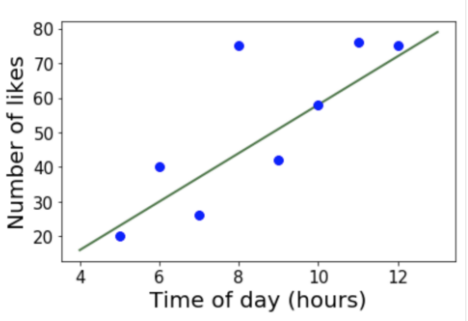
Linear and Logistic Regression

Goal: Predict numerical outputs from previously unseen inputs.

Linear regression:

Models: Linear relationship between inputs and outputs.

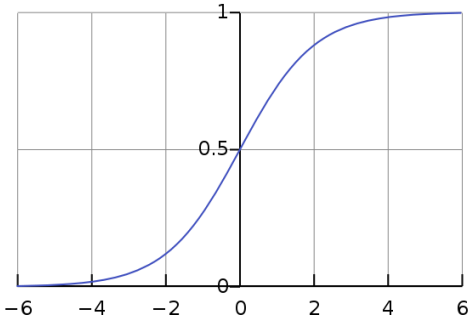
Find the line ($\hat{y} = mx + b$) that minimizes error.



Logistic regression:

Models: Sigmoidal relationship between inputs and outputs.

Find the model ($\hat{y} = \frac{1}{1+e^{-(w_1 x + w\_2)}}$) that minimizes the error



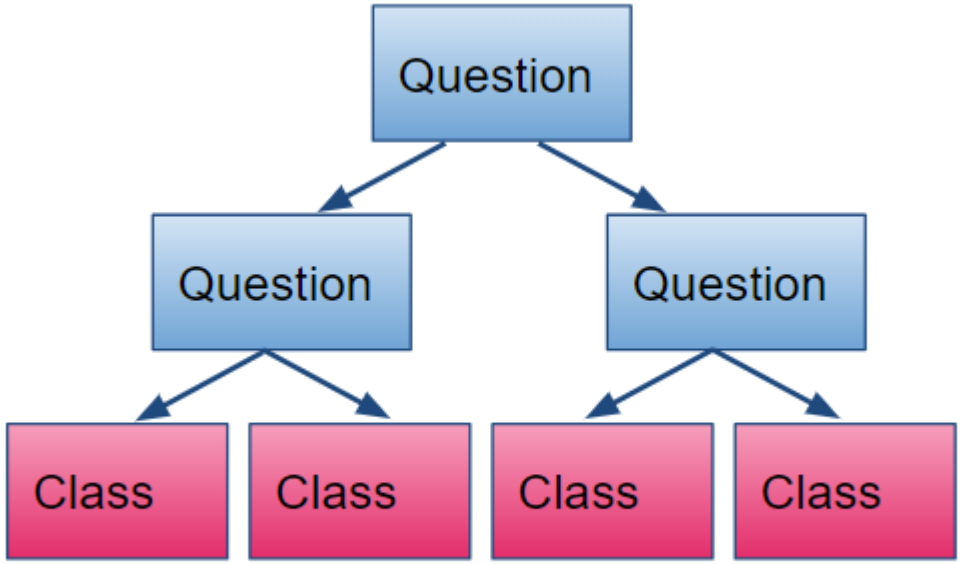
Decision Trees

Goal: Classification.

Decision Trees

Goal: Classification.

Method: Create tree of questions to determine classification.



Decision Trees

Goal: Classification.

Method: Create tree of questions to determine classification.

Method: Constructing the Tree:

- Select “best split” feature.
- Create new node based on this feature.
- Separate training dinner based on split.
- Repeat until:
  - *Group’s examples have same type.*
  - *Group is too small.*
  - *Tree is too large.*
- Create child node representing classification.

Decision Trees

Goal: Classification.

Method: Create tree of questions to determine classification.

Method: Constructing the Tree:

- Select “best split” feature: **Maximum information gain.**
- Create new node based on this feature.
- Separate training dinner based on split.
- Repeat until:
  - *Group’s examples have same type.*
  - *Group is too small.*
  - *Tree is too large.*
- Create child node representing classification.

Nearest Neighbors

Goal: Classification.

Nearest Neighbors

Goal: Classification.

Idea: Classify similar nodes similarly.

Nearest neighbors, k-nearest neighbors.

Naïve Bayes

Goal: Classification.

Naïve Bayes

Goal: Classification.

Assumption: Independence.

Naïve Bayes

Goal: Classification.

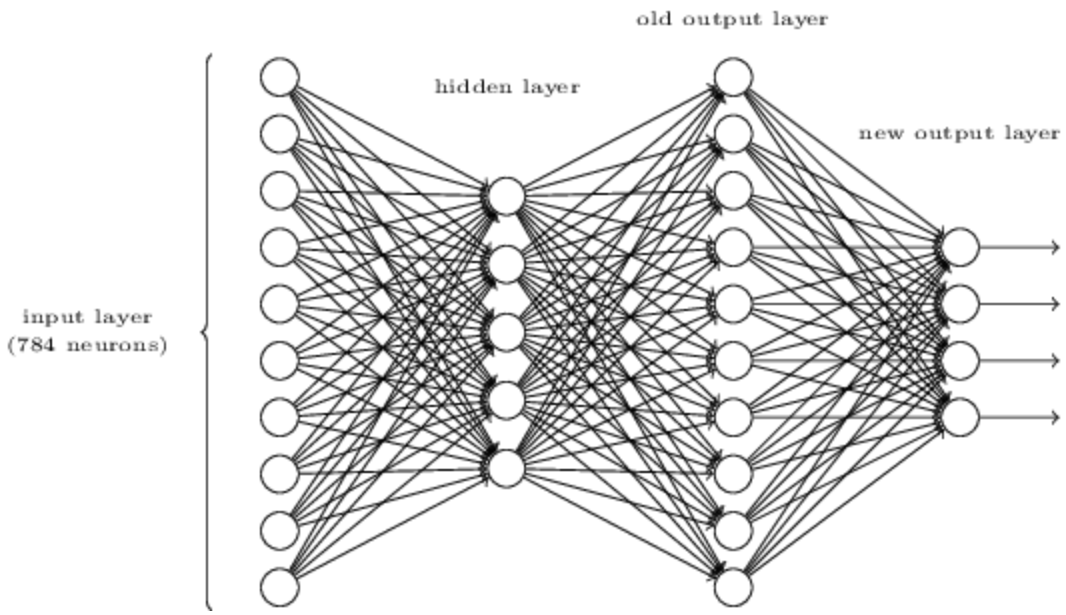
Assumption: Independence.

Method: Use Bayes' Rule to find probability of each class, and choose class with highest probability.

Bayes' Rule: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Neural Networks

Structure:



Training: Backward propagation.

K-Means Clustering

Unsupervised Learning.

K-Means Clustering

Unsupervised Learning.

Method:

- Pick k random centroids.
- Assign each example to the closest centroid.
- Recompute the centroids.

Q&A

Intro to Python and CS Concepts

Intro to Machine Learning

Data Exploration and Stats

Linear and Logistic Regression

Decision Trees and Nearest Neighbors

Probability and Naïve Bayes

Unsupervised Learning

Neural Networks

Future of AI