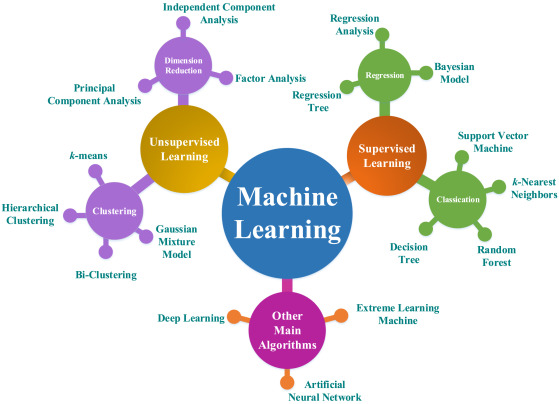
Machine Learning Cheat Sheet

Mackenzie Biduk

Diagram

Description automatically generated



# Linear Algebra

|  |  |
| --- | --- |
| **Linear Algebra** | |
| **Matrix Multiplication** | **Transpose Matrix** |
| x = =  [3x2] x [2x1] = [3x1] |  |
| **Determinant (2x2)** | **Inverse Matrix (2x2)** |
|  |  |
| **Determinant (3x3)** | **Inverse Matrix (3x3)** |
| **https://ardoris.wordpress.com/2008/07/18/general-formula-for-the-inverse-of-a-3x3-matrix/** | <https://www.wikihow.com/Find-the-Inverse-of-a-3x3-Matrix>  A picture containing text, clock, gauge  Description automatically generated  Diagram, text  Description automatically generated with medium confidence |

# Optimization Algorithms

**Optimization Algorithms**

* Gradient Descent
* Conjugate Gradient (Faster, no , complex) \*Use libraries
* BFGS (Faster, no , complex) \*Use libraries
* L-BFGS (Faster, no , complex) \*Use libraries

# Types of Machine Learning

**Types of Machine Learning**

1. [Supervised Learning](#_Supervised_Learning)
   1. [Regression](#_Regression)
   2. [Classification](#_Classification)
   3. Neural Networks
2. [Unsupervised Learning](#_Unsupervised_Learning)
3. [Reinforcement Learning](#_Reinforced_Learning)
4. [Recommender System](#_Recommender_Systems)

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Definition   * Given the “right answers” * Known relation between the variables of the data and the output   + Given X   + Given Y | Definition   * Unknown “answers” * Unknown relation to variables from the data   + Given X   + Unknown Y |
| Models   * Regression * Classification | Models   * Clustering * Non-Clustering |
| Problems   * Predicting Housing Prices * Predicting Cancerous or Malignant * Predicting known Categories | Problems   * Organizing Computer Clusters * Social Network Analysis * Market Segmentation * Astronomical Analysis |

|  |  |
| --- | --- |
| **Reinforcement Learning** | **Recommender Systems** |
| Definition | Definition |
| Models | Models |
| Problems | Problems |

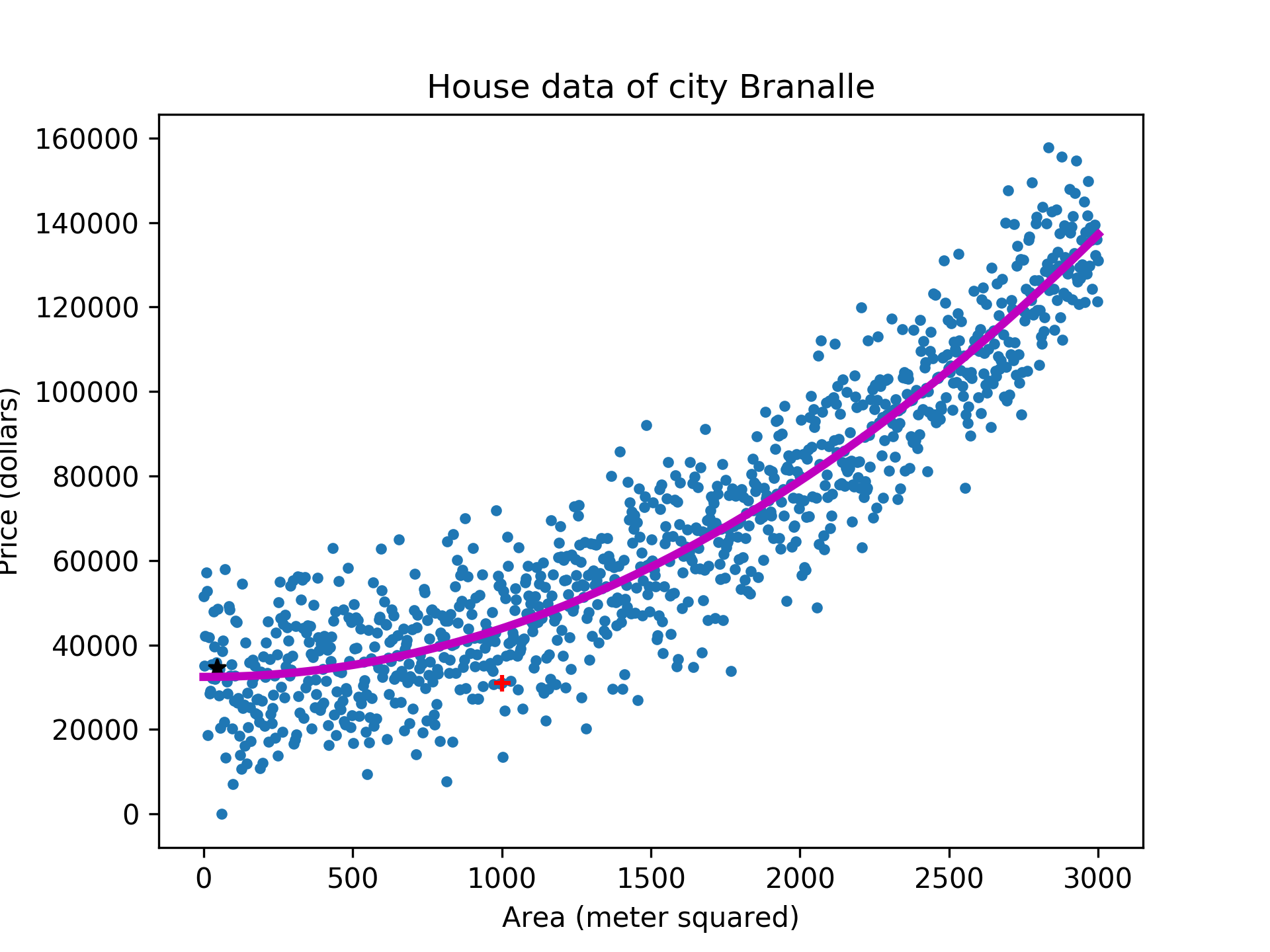
## Supervised Learning

|  |
| --- |
| **Supervised Learning** |
| Definition   * Given the “right answers” * Known relation between the variables of the data and the output   + Given X   + Given Y |
| Models   * Regression * Classification * Neural Networks |
| Problems   * Predicting Housing Prices * Predicting Cancerous or Malignant * Predicting known Categories |

|  |  |
| --- | --- |
| **Supervised Learning** | |
| **Regression** | **Classification** |
| Definition   * Continuous Outputs | Definition   * Discrete Outputs |
| Problems   * Predicting Prices | Problems   * Yes or No * Categories |

### Regression

Regression uses pre-existing data with one or more predictor variables (x). These predictor variables are used to predict a continuous output (y) or line-of-best-fit. From here, you may predict an output y(i) based on any given predictor variables x(i).



#### Cost Function J(ϴ)

|  |  |  |
| --- | --- | --- |
| **Regression - Cost Functions** | | |
| Cost Function | Regularized Cost Function |  |
|  |  |  |
|  |  |  |
| |  |  |  |  | | --- | --- | --- | --- | | xj(i) | Feature j of the ith Training Example | nj | # Features | | y | Target | hϴ(x) | Projected Target | | (x(i),y(i)) | ith Training Example | J(ϴ) | Cost Function (Squared Error Function) | | m | # Training Examples | ϴj | Theta for feature j | |  | Regularization Parameter |  | | | |

#### Hypothesis hϴ(x)

|  |  |  |
| --- | --- | --- |
| **Regression – Hypothesis** | | |
| Linear Regression | Non-Linear Regression | Multivariate Regression |
|  |  | Additional Features ... |

#### Minimizing J(ϴ)

|  |  |  |
| --- | --- | --- |
| **Optimizing Techniques** | | |
| Gradient Descent | Normal Equation | Normal Regularized Equation |
| **⢰**  **⢰** | Add a column of 1s to X to represent x0.  ⟶  hϴ(x) = x | Add a column of 1s to X to represent x0.  ⟶  hϴ(x) = x |
| Need  Need iterations  Works well for n>> | No  No iterations  Slow for n>104 | |
| |  |  | | --- | --- | | 𝛼 | Learning Rate (Step Size) | | | |

##### Gradient Descent

|  |  |  |  |
| --- | --- | --- | --- |
| **Gradient Descent** |  | **Improving Gradient Descent**  \*Pick Only One\* | |
| 1. Choose a set of initial parameters for 2. Choose a step size    * too small ⟶ slow convergence    * too large ⟶ overshoot or divergence 3. Converge to global minimum (convex).   **Diagram  Description automatically generated**   * J(ϴ)MIN should decrease for every iteration * J(ϴ) MIN should converge given enough iterations   A picture containing chart  Description automatically generated |  | Feature Scaling | Feature Scaling |
| Diagram  Description automatically generated | Diagram  Description automatically generated |
|  | |  |  | | --- | --- | |  | Mean of feature | | |

##### Normal Equation

|  |  |
| --- | --- |
| Normal Equation | |
| No Regularization | Regularization |
| * + - 1. Add a column of 1s to X to represent x0.   ⟶   * + - 1. hϴ(x) = x   If is non-invertible   * Redundant Features * Too many features   + Delete Features   + Regularization | * + - 1. Add a column of 1s to X to represent x0.   ⟶   * + - 1. .       2. hϴ(x) = x *.*   If λ>0, the normal regularized equation is always invertible. |

### Classification

Classification is a predictive model where a class label is predicted for a given example of input data.

Chart, scatter chart

Description automatically generated

#### Cost Function J(ϴ)

|  |  |
| --- | --- |
| **Cost Function** | |
| Logistic Regression | Regularized Logistic Regression |
|  |  |

#### Hypothesis hϴ(x)

|  |
| --- |
| **Classification – Hypothesis** |
|  |
| |  |  | | --- | --- | | g(z) | Sigmoid Function or Logic Function | |

#### Minimizing J(ϴ)

|  |  |  |
| --- | --- | --- |
| **Optimization Techniques** | | |
| Gradient Descent | Normal Equation | Normal Equation Regularization |
| 1. Choose a set of initial parameters for 2. Choose a step size    * too small ⟶ slow convergence    * too large ⟶ overshoot or divergence 3. Converge to global minimum (convex).   **Diagram  Description automatically generated**   * J(ϴ)MIN should decrease for every iteration * J(ϴ) MIN should converge given enough iterations   A picture containing chart  Description automatically generated | * + - 1. Add a column of 1s to X to represent x0.   ⟶   * + - 1. hϴ(x) = x   If is non-invertible   * Redundant Features * Too many features   + Delete Features   + Regularization | 1. Add a column of 1s to X to represent x0.   ⟶   1. . 2. hϴ(x) = x *.*   If λ>0, the normal regularized equation is always invertible. |
| Need  Need iterations  Works well for n>> | No  No iterations  Slow for n>104 | |

##### Gradient Descent

|  |  |  |  |
| --- | --- | --- | --- |
| **Gradient Descent** |  | **Improving Gradient Descent**  \*Pick Only One\* | |
|  |  | Feature Scaling | Feature Scaling |
| **⢰**  **⢰** | Diagram  Description automatically generated | Diagram  Description automatically generated |
| |  |  | | --- | --- | | 𝛼 | Learning Rate (Step Size) | |  | |  |  | | --- | --- | |  | Mean of feature | | |

##### Normal Equation

|  |  |
| --- | --- |
| **Normal Equation** | |
| Normal Equation | Regularized Normal Equation |
| Add a column of 1s to X to represent x0.  ⟶  hϴ(x) = x | Add a column of 1s to X to represent x0.  ⟶  hϴ(x) = x |
| |  |  | | --- | --- | | 𝛼 | Learning Rate (Step Size) | | |

### Neural Networks

Diagram

Description automatically generated

Diagram, engineering drawing

Description automatically generated

## Unsupervised Learning

## Reinforced Learning

## Recommender Systems

# Bias and Variance

*Error between the average model prediction and the true value*

A picture containing chart

Description automatically generated

|  |  |
| --- | --- |
| **High Bias** | **High Variance** |
| Underfitting  Scatter chart  Description automatically generated  High Training Error  High Test Error  Solution   * Regularization   + Increase magnitude of ϴj   + Decrease λ * Increase number of features | Overfitting  A picture containing chart  Description automatically generated  Low Training Error  High Test Error  Solution   * Regularization   + Reduce magnitude of ϴj   + Increase λ * Reduce the number of features   + Manually chose features to keep   + Model Selection Algorithm |