**Cheat Sheet: Supervised and Unsupervised Learning**

**Supervised Learning**

* **Definition**: A type of machine learning where the model is trained on labeled data.
* **Goal**: To learn the mapping from input features to the target output.
* **Common Algorithms**:
  + **Linear Regression**
  + **Logistic Regression**
  + **Decision Trees**
  + **Support Vector Machines (SVM)**
  + **Neural Networks**

**Unsupervised Learning**

* **Definition**: A type of machine learning where the model is trained on unlabeled data.
* **Goal**: To find hidden patterns or intrinsic structures in the input data.
* **Common Algorithms**:
  + **K-Means Clustering**
  + **Hierarchical Clustering**
  + **Principal Component Analysis (PCA)**
  + **Association Rules (e.g., Apriori algorithm)**
  + **Autoencoders**

**Linear Regression**

* **Definition**: A supervised learning algorithm used to predict a continuous target variable based on one or more input features.
* **Formula**: y=β0+β1x1+β2x2+…+βnxny = \beta\_0 + \beta\_1x\_1 + \beta\_2x\_2 + \ldots + \beta\_nx\_ny=β0​+β1​x1​+β2​x2​+…+βn​xn​
  + yyy: Target variable
  + x1,x2,…,xnx\_1, x\_2, \ldots, x\_nx1​,x2​,…,xn​: Input features
  + β0\beta\_0β0​: Intercept
  + β1,β2,…,βn\beta\_1, \beta\_2, \ldots, \beta\_nβ1​,β2​,…,βn​: Coefficients

**Example**

Suppose you want to predict the price of a house based on its size.

* **Input Feature (x)**: Size of the house in square feet
* **Target Variable (y)**: Price of the house
* **Linear Regression Equation**: Price=β0+β1×Size\text{Price} = \beta\_0 + \beta\_1 \times \text{Size}Price=β0​+β1​×Size

**Cost Function**

* **Definition**: A function that measures the performance of a machine learning model for given data. It quantifies the error between predicted values and the actual values.
* **Common Cost Function for Linear Regression**: Mean Squared Error (MSE)
  + **Formula**: MSE=1n∑i=1n(y^i−yi)2\text{MSE} = \frac{1}{n} \sum\_{i=1}^n (\hat{y}\_i - y\_i)^2MSE=n1​∑i=1n​(y^​i​−yi​)2
    - nnn: Number of data points
    - y^i\hat{y}\_iy^​i​: Predicted value for the i-th data point
    - yiy\_iyi​: Actual value for the i-th data point

**Gradient Descent**

* **Definition**: An optimization algorithm used to minimize the cost function by iteratively adjusting the model parameters.
* **Process**:
  1. **Initialize** the parameters (e.g., β0,β1,…,βn\beta\_0, \beta\_1, \ldots, \beta\_nβ0​,β1​,…,βn​).
  2. **Compute the gradient** of the cost function with respect to each parameter.
  3. **Update the parameters** in the direction opposite to the gradient to reduce the cost.
  4. **Repeat** steps 2 and 3 until the cost function converges to a minimum.

**Gradient Descent Formula**

* **Update Rule**: βj:=βj−α∂J∂βj\beta\_j := \beta\_j - \alpha \frac{\partial J}{\partial \beta\_j}βj​:=βj​−α∂βj​∂J​
  + βj\beta\_jβj​: Parameter to be updated
  + α\alphaα: Learning rate (step size)
  + ∂J∂βj\frac{\partial J}{\partial \beta\_j}∂βj​∂J​: Partial derivative of the cost function JJJ with respect to βj\beta\_jβj​

**Convergence**

* **Convergence**: The algorithm is said to converge when the changes in the cost function become very small, indicating that the optimal parameters have been found.
* **Relation to Gradient**: As the parameters get updated, the gradient approaches zero, indicating that the minimum of the cost function has been reached.

**Summary**

* **Supervised Learning**: Uses labeled data to predict outcomes.
* **Unsupervised Learning**: Finds patterns in unlabeled data.
* **Linear Regression**: Predicts continuous values; uses a linear equation.
* **Cost Function**: Measures model error; commonly MSE for linear regression.
* **Gradient Descent**: Optimizes parameters to minimize the cost function; iterates until convergence.

This cheat sheet provides a concise overview of key concepts in supervised and unsupervised learning, focusing on linear regression and optimization techniques.

### Interview Questions on Supervised and Unsupervised Learning

#### Supervised Learning

1. **What is supervised learning?**
   * **Answer**: Supervised learning is a type of machine learning where the model is trained on labeled data. The goal is to learn the mapping from input features to the target output.
2. **Can you name some common algorithms used in supervised learning?**
   * **Answer**: Common supervised learning algorithms include Linear Regression, Logistic Regression, Decision Trees, Support Vector Machines (SVM), and Neural Networks.
3. **What are the main differences between supervised and unsupervised learning?**
   * **Answer**: Supervised learning uses labeled data to train the model and predict outcomes, while unsupervised learning uses unlabeled data to find hidden patterns or structures. Supervised learning aims to learn a mapping from inputs to outputs, whereas unsupervised learning focuses on clustering, association, or dimensionality reduction.

#### Unsupervised Learning

1. **What is unsupervised learning?**
   * **Answer**: Unsupervised learning is a type of machine learning where the model is trained on unlabeled data. The goal is to find hidden patterns or intrinsic structures in the input data.
2. **Can you name some common algorithms used in unsupervised learning?**
   * **Answer**: Common unsupervised learning algorithms include K-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), Association Rules (e.g., Apriori algorithm), and Autoencoders.
3. **What are some applications of unsupervised learning?**
   * **Answer**: Applications of unsupervised learning include customer segmentation, anomaly detection, data compression, market basket analysis, and image segmentation.

#### Linear Regression

1. **What is linear regression?**
   * **Answer**: Linear regression is a supervised learning algorithm used to predict a continuous target variable based on one or more input features. It models the relationship between the target and the input features using a linear equation.
2. **Can you explain the equation of linear regression?**
   * **Answer**: The linear regression equation is y=β0+β1x1+β2x2+…+βnxny = \beta\_0 + \beta\_1x\_1 + \beta\_2x\_2 + \ldots + \beta\_nx\_ny=β0​+β1​x1​+β2​x2​+…+βn​xn​, where yyy is the target variable, x1,x2,…,xnx\_1, x\_2, \ldots, x\_nx1​,x2​,…,xn​ are the input features, β0\beta\_0β0​ is the intercept, and β1,β2,…,βn\beta\_1, \beta\_2, \ldots, \beta\_nβ1​,β2​,…,βn​ are the coefficients.
3. **What are some assumptions of linear regression?**
   * **Answer**: Assumptions of linear regression include linearity, independence of errors, homoscedasticity (constant variance of errors), and normally distributed errors.

#### Cost Function

1. **What is a cost function in the context of linear regression?**
   * **Answer**: In linear regression, the cost function measures the error between the predicted values and the actual values. It quantifies the performance of the model. A common cost function is Mean Squared Error (MSE).
2. **Can you explain the Mean Squared Error (MSE) formula?**
   * **Answer**: The MSE formula is MSE=1n∑i=1n(y^i−yi)2\text{MSE} = \frac{1}{n} \sum\_{i=1}^n (\hat{y}\_i - y\_i)^2MSE=n1​∑i=1n​(y^​i​−yi​)2, where nnn is the number of data points, y^i\hat{y}\_iy^​i​ is the predicted value for the i-th data point, and yiy\_iyi​ is the actual value for the i-th data point.

#### Gradient Descent

1. **What is gradient descent?**
   * **Answer**: Gradient descent is an optimization algorithm used to minimize the cost function by iteratively adjusting the model parameters. It updates the parameters in the direction opposite to the gradient of the cost function.
2. **What is the significance of the learning rate in gradient descent?**
   * **Answer**: The learning rate determines the step size for each iteration of the gradient descent. If the learning rate is too high, the algorithm may overshoot the minimum. If it is too low, the algorithm may converge too slowly.
3. **How do you know when gradient descent has converged?**
   * **Answer**: Gradient descent is said to converge when the changes in the cost function become very small, indicating that the optimal parameters have been found. This is typically indicated by the gradient approaching zero.