Gretel’s Quick and Dirty Guide to Supervised Machine Learning

# Is the Target a **Number** (like housing price)? 🡪 Regression model

* **Linear Regression -** Finds linear equation with best fit based on R2 value
* **Ridge -** Same as linear regression, but with additional parameter to prevent overfitting
* **Lasso -** Same as ridge regression, but with different method to prevent overfitting
* **ElasticNet -** Balance between Lasso and Ridge methods of preventing overfitting

# Is the Target a **Category** (like True/False, or type of car)? 🡪 Classification model

* **Logistic Regression -** Finds equation to determine probability of target result given input features
* **Decision Tree -** Determines category by using T/F questions to navigate through all features
* **Random Forest -** Create multiple small decision trees from randomly selected subset of features to find how to best weigh the features to optimize model; good at preventing overfitting.
  + Can use RF to get features importance.
* **K Nearest Neighbors -** Determines classification of new data by looking at training data and finding the majority of classifications of the k nearest neighbors.
  + Can also use to get features importance
  + Can use to impute missing data
* **Support Vector Machine -** Finds best boundary (“optimal hyperplane”) to separate data of different classes.
  + *Kernel* refers to the type of equation you are going to use to define your hyperplane
  + *C* refers to how wide the margins are to the boundaries (big C 🡪 small margin)
    - Large margin finds the optimal boundary; small margin minimizes classification errors
    - <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>

## Documentation Links

* **Linear Regression:** <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>
* **Ridge:** <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html#sklearn.linear_model.Ridge>
* **Lasso:** <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html#sklearn.linear_model.Lasso>
* **ElasticNet:** <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html#sklearn.linear_model.ElasticNet>
* **Logistic Regression:** <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression>
* **Decision Tree:** <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier>
* **Random Forest:** <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier>
* **K Nearest Neighbors:** <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier>
* **Support Vector Machine:** <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>

*If you want to get deeper into the math behind these models, go to Gretel’s favorite YouTube Channel:*

[*https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw*](https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw)