# Scikit-Learn

**Package: `sklearn.linear_model`**

| Linear Regression | Logistic Regression | Class/Function(s) | Description |
|---|---|---|---|
| ✓ | - | `LinearRegression(fit_intercept=True)` | Returns an ordinary least squares Linear Regression model. |
| - | ✓ | `LogisticRegression( fit_intercept=True, penalty='l2', C=1.0)` | Returns an ordinary least squares Linear Regression model. Hyperparameter C is inverse of regularization parameter, C = 1/λ. |
| ✓ | - | `Lasso()`, `Ridge()` | Returns a Lasso (L1 Regularization) or Ridge (L2 regularization) linear model, respectively. |
| ✓ | ✓ | `model.fit(X, y)` | Fits the scikit-learn `model` to the provided `X` and `y`. |
| ✓ | ✓ | `model.predict(X)` | Returns predictions for the `X` passed in according to the fitted `model`. |
| - | ✓ | `model.predict_proba(X)` | Returns predicted probabilities for `X` passed in according to the fitted `model`. If binary classes, returns probabilities for both classes 0 and 1. |
| ✓ | ✓ | `model.coef_` | Estimated coefficients for the linear model, not including the intercept. |
| ✓ | ✓ | `model.intercept_` | Bias/intercept term of the linear model. Set to 0.0 if `fit_intercept=False`. |

**Package: `sklearn.model_selection`**

| | | Function | Description |
|---|---|---|---|
| | | `train_test_split(*arrays, test_size=0.2)` | Returns two random subsets of each array passed in, with 0.8 of the array in the first subset and 0.2 in the second subset. |

# Probability

Let $X$ have a discrete probability distribution $P(X = x)$. $X$ has expectation $\mathbb{E}[X] = \sum_x xP(X = x)$ over all possible values $x$, variance $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$, and standard deviation $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$.

The covariance of two random variables $X$ and $Y$ is $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. If $X$ and $Y$ are independent, then $\mathrm{Cov}(X, Y) = 0$.

| Notes | Property of Expectation | Property of Variance |
|---|---|---|
| $X$ is a random variable. | | $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ |
| $X$ is a random variable, $a, b \in \mathbb{R}$ are scalars. | $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ | $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ |
| $X, Y$ are random variables. | $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ | $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$ |
| $X$ is a Bernoulli random variable that takes on value 1 with probability $p$ and 0 otherwise. | $\mathbb{E}[X] = p$ | $\mathrm{Var}(X) = p(1 - p)$ |

### Central Limit Theorem

Let $(X_1, \ldots, X_n)$ be a sample of independent and identically distributed random variables drawn from a population with mean $\mu$ and standard deviation $\sigma$. The sample mean $\overline{X}_n = \sum_{i=1}^{n} X_i$ is normally distributed, where $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathrm{SD}(\overline{X}_n) = \sigma/\sqrt{n}$.

### Model Risk

Suppose for each individual with fixed input $x$, we observe a random response $Y = g(x) + \epsilon$, where $g$ is the true relationship and $\epsilon$ is random noise with zero mean and variance $\sigma^2$.

For a new individual with fixed input $x$, define our random prediction $\hat{Y}(x)$ based on a model fit to our observed sample $(\mathbb{X}, \mathbb{Y})$. The model risk is the mean squared prediction error between $Y$ and $\hat{Y}(x)$: $\mathbb{E}[(Y - \hat{Y}(x))^2] = \sigma^2 + \left(\mathbb{E}[\hat{Y}(x)] - g(x)\right)^2 + \mathrm{Var}(\hat{Y}(x))$.