Unit 4 Glossary

Dot product

Also referred to as scalar product, a special operation that takes the element's product of each vector and then sums them together. It is typically represented with what looks like a period or a dot floating between the two vectors.

Gradient descent

A numerical optimization algorithm used to train and optimize a logistic regression model using loss functions. Gradient descent iteratively updates the model parameters until a loss function is minimized.

Hessian matrix

A square matrix of second-order partial derivatives of a scalar-valued function, or scalar field.

Identity matrix

A special matrix whose diagonal elements are one, and all other elements are zero.

Intercept

Sometimes called the "constant." Intercept in a linear regression model is the mean value of the response variable when all of the predictor variables in the model are equal to zero. Intercept in a logistic regression is the "log odds" of the response variable, not the mean.

Inverse Logit

One of three steps in the logistic regression model. The inverse logit step transforms the output of the linear step into a probability prediction $P(y|X)$ between 0-1.

Learning rate

A common logistic regression hyperparameter (also commonly known as the step size), learning rate dictates the speed of gradient descent. The ideal learning rate is one that reaches global minima in a fast and efficient manner.

Linear models

A class of supervised learning models that are represented by an equation: simple to implement, fast to train, and have lower complexity. Logistic regression is an example of a linear model. In a linear model, the form of a model must fit a very specific format.

Linear regression

A supervised machine learning algorithm used for regression problems. Linear regression finds a linear relationship between one or more features and a label (such as a price or an age.) There are two types of linear regression models: Simple linear regression finds the linear relationship between one feature and one label, and multiple linear regression finds the linear relationship between multiple features and one label.

Linear step

One of three steps in the Logistic Regression model. The linear step computes a value $z(X)$ by taking the linear sum of feature values X with model weights W and an intercept term α (also known as bias, β, in some literature).

Log loss

A popular loss function used to measure the performance of a some classification models. This loss function is used for both logistic regression and neural networks.

Logistic regression

A linear classification method that is trained by iteratively tuning a set of weights to minimize the log loss.

Loss function

Specialized mathematical function that represents how well our models predict the labels. A loss function quantifies the amount of error a model makes against the training dataset.

Matrix inverse

A matrix that when multiplied by the original matrix produces the identity matrix.

Matrix multiplication

A binary operation that produces a matrix from two matrices.

Mean squared error

Loss function commonly used to measure the performance of a regression model such as linear regression.

Non-linear models

The opposite of linear models, non-linear models can take many different forms: Non-linear models have more complexity, meaning they can draw more sophisticated curves to fit arbitrary patterns in the data.

NumPy

A Python library that adds support for multi-dimensional arrays and matrices. NumPy also provides an extensive collection of high-level mathematical functions to do element-wise operations on entire arrays.

Optimization algorithm

An algorithm that uses a loss function to evaluate a model's loss and then adjusts the model parameters accordingly to reduce loss. It continues this process until an optimal model is produced. A popular optimization algorithm is gradient descent.

Ordinary least squares (OLS)

A non-iterative method used in linear regression to minimize the sum of the squared errors between the model predictions and the actual values.

Overfitting

A model failure mode that occurs when a model is too complex. It learns the training data so closely that it does not generalize well to new data. An overfit model has low training error but poor generalization.

Regularization

The penalty on a model's complexity; helps prevent overfitting. Different kinds of regularization include L1 regularization and L2 regularization.

Scikit-learn

Software for Python that has a very wide range of algorithmic options, covering regression, classification, and unsupervised learning. It also provides rich libraries for data preparation, model selection, and evaluation.

Zero-one loss

A simple loss function that counts how many mistakes an hypothesis function h makes on the training set. This loss function is not suited for use in the training process to optimize the model but is suited for the evaluation of classification models.