General Assembly

STUDY GUIDE

VECTORS IN PYTHON

Key Terms & Definitions

- A **scalar** is a single number that represents **magnitude**. Scalars are often used for measuring things such as length, distance, and speed. A scalar is typically written as a non-bold, lowercase character.
- A **vector** is an ordered list of scalars representing **magnitude and direction**. (This is called the "component form.") In linear algebra, if you see a list of numbers enclosed in parentheses or brackets, you can assume that it's a vector. A vector is typically depicted either as a bold, lowercase letter (**a**, for example) or a lowercase letter with an arrow on top.
- The process of converting non-numeric data points to vectors is called **vectorizing** the data.
- In scikit-learn, any function that includes the word **vectorizer** takes non-numeric input and returns numeric vectors as its output.
- Using n scalars, one binary scalar per category is called a **one hot encoding**.
- **Coefficients** are indicators of how steep or extreme a gradient or value is. Most often, this indicates the slope of a line. These are often denoted by capital letters.
- **Parametric** models rely on a mathematical formula with unknown coefficients/parameters such as linear regression, logistic regression, and neural networks.
- **Regression** is the process of using a line to either separate two groups or predict an exact value along that line.
  - In two dimensions, we solve for a line that separates the points ($y = Ax + B$).
  - In three dimensions, we solve for the equation of a **plane** ($z = Ax + By + C$).
  - In more than three dimensions, we solve for a **hyperplane** ($w = Ax + By + Cz + ...$).
- Polynomial **degrees** are the largest powers to which the variables are raised. In a **linear equation**, each term is only one variable raised to (at most) a power of one (i.e., the polynomial has a **degree** of one).
- **Cosine similarity** compares two vectors without regard to their magnitude, focusing only on direction.
- **Orthogonal** means at a 90-degree angle.
- **High-dimensional data** means that each point is represented by a long list of ordered numbers.
- **Normalizing a vector** finds the vector with a magnitude of one in the same direction as the original vector. This is called a **unit vector**, as the vector has unit length.

    - This is found by locating the point at which a vector hits along a **unit circle**, which is a circle with a radius of one that surrounds the origin of the vector. A unit circle in three dimensions is a **unit sphere**. In higher dimensions, it's a **unit hypersphere**.
- **Scaling a vector** refers to multiplying a vector by a scalar, changing the vector's magnitude but not its direction.
- A **unit circle** in three dimensions is a **unit sphere**. In higher dimensions, it's a **unit hypersphere**.

Guiding Questions

1) What is the benefit of normalizing a vector? 2) Describe how to compare two vectors to one another. 3) Describe a situation in which **one hot encoding** would be useful. 4) In what situations would you want to compare the magnitude of a set of vectors? The direction? Both?

Additional Resources

1. Khan Academy's Vector Unit
2. Math Is Fun's Vector Cheat Sheet
3. Connecting Vectors to Matrices for Data Science