Feature Engineering

Papangkorn Inkeaw, Ph.D.

Feature Construction

Chapter 4 (Part II)

Feature Expansion

- A single numeric predictor can be expanded to many predictors.
- These one-to-many transformations of the data can be used to improve model performance.
- For example, in cases where the target follows a quadratic relationship with a variable, creating a second-degree polynomial of the feature allows us to use it in a linear model.

Polynomial Expansion

- A basis expansion of a predictor x can be achieved by deriving a set of functions $f_i(x)$ that can be combined using a linear combination.

$$f(x) = \sum_i \beta f_i(x)$$

- A basic function is a polynomial function.
- Polynomial expansion serves to
  - Automate the creation of new features
  - Capture feature interaction
  - Capture potential non-linear relationships between the original variables and the target.

Polynomial Expansion

- Polynomial expansion of a predictor x with a degree d:

$$[1, x, x^2, \dots, x^d]$$

- Polynomial expansion of a combination of $x_1, x_2, \dots, x_p$ with the second degree:

$$[x_1, x_2, \dots, x_p]^2 = [x_1, x_2, \dots, x_p][x_1, x_2, \dots, x_p] = [1, x_1, x_2, .., x_p. x_1^2, x_1x_2, x_1x_3, \dots, x_2^2, x_2x_3, \dots, x_p^2\ ]$$

- Polynomial expansion of a combination of $x_1, x_2, \dots, x_p$ with the third degree:

$$[x_1, x_2, \dots, x_p]^3 = [x_1, x_2, \dots, x_p]^2[x_1, x_2, \dots, x_p]$$

Polynomial Expansion

- The “degree” of the polynomial is used to control the number of features added.
- Generally speaking, it is unusual to use d greater than 3 or 4 because for large values of d, the polynomial curve can become overly flexible and can take on some very strange shapes.

Polynomial Expansion

Example

Polynomial expansion of $[a, b, c]$ with the two degree:

$$[a, b, c]^2 = [1, a, b, c, a^2, ab, ac, b^2, bc, c^2]$$

Polynomial expansion of $[a, b, c]$ with the third degree:

$$[a, b, c]^3 = [1, a, b, c, a^2, ab, ac, b^2, bc, c^2][a, b, c]$$
$$= [1, a, b, c, a^2, ab, ac, b^2, bc, c^2, a^3, a^2b, a^2c, ab^2, abc, ac^2, b^3, b^2c, bc^2, c^3]$$

References & Study Resources

- Soledad Galli. (2020). *Python Feature Engineering Cookbook*. Packt Publishing.
- Max Kuhn and Kjell Johnson. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.