Math 342W/642/742W

Recitation – Day #13 (3.27.25)

I. More on Logistic Regression

(i) Which random variable is used to model Y_i?

$$Y_i \sim \text{Bern}\left(\Phi(\boldsymbol{x} \cdot \boldsymbol{\beta})\right)$$

(ii) What is g_{pr_0}, the default mode for $P(Y_i = 1)$?

$$g_{pr_0} = \bar{y} \quad (\textit{the proportion of } y\textit{'s that are 1})$$

(iii) Derive "*log-odds*" from the logistic model.

$$\begin{aligned}
\hat{p}(\boldsymbol{x}) = g_{pr}(\boldsymbol{x}) &= \frac{1}{1+e^{-\boldsymbol{b}\cdot\boldsymbol{x}}} \\
&\implies \frac{1}{\hat{p}} = 1 + e^{-\boldsymbol{b}\cdot\boldsymbol{x}} \\
&\implies 1 - \frac{1}{\hat{p}} = e^{-\boldsymbol{b}\cdot\boldsymbol{x}} \\
\textit{odds-against} &\implies \frac{\hat{p}-1}{\hat{p}} = e^{-\boldsymbol{b}\cdot\boldsymbol{x}} \\
\textit{odds} \in (0,\infty) &\implies \frac{\hat{p}}{1-\hat{p}} = e^{\boldsymbol{b}\cdot\boldsymbol{x}} \\
\textit{log-odds} \in (-\infty,\infty) = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) &= \boldsymbol{b}\cdot\boldsymbol{x}
\end{aligned}$$

(iv) Define the two "*proper* scoring rules" for probability estimation:

We use a "proper scoring rule" such that

$$\forall i \quad f_{pr}(\boldsymbol{x}_i) = \text{argmax}\{S(\hat{p}_i, y_i)\}$$

We'll consider two scoring rules:

(1) Brier Score

$$s_i = -(y_i - \hat{p}_i)^2 \leq 0$$

(2) Log-Scoring Rule

$$s_i = y_i \ln(\hat{p}_i) + (1-y_i)\ln(1-\hat{p}_i) \leq 0$$

We then consider: $\frac{1}{n}\sum s_i = \bar{s}$

II. Polynomial Modeling

(i) What type of error does transforming raw features into exponentiated values help reduce?

Misspecification error is reduced

(ii) What are ***first-order*** interactions?

$$\text{e.g.,}\quad \hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 \underbrace{x_1 \cdot x_2}_{\text{first-order}}$$

(iii) What is ***Weierstrauss' Theorem***?

Every continuous function defined on a closed interval $[a, b]$ can be uniformly approximated as closely as desired by a polynomial function.

(iv) What is the distinction between p_{raw} and p?

- $p_{raw} = \#$ of original features
- $p = \#$ of resulting features after adding derived features, transformations, interactions, etc.

(v) What is the candidate set $\mathcal{H}$ when modeling with transformed (exponentiated) features?

e.g., when $p_{raw} = 1 \implies p = 2$

$$\mathcal{H} = \{w_0 + w_1 x + w_2 x^2 \mid w_0, w_1, w_2 \in \mathbb{R}\}$$

(vi) What is the matrix X associated with fitting a polynomial to a set of n points?

The Vandermonde Matrix, X, has full rank and $\det(X) = \prod_{i=1}^{n} \prod_{j=1}^{n} (x_j - x_i) \neq 0$ if $x_1, \ldots, x_n$ are distinct values.

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \ldots & x_n^{n-1} \end{bmatrix}$$

(vii) What is ***Runge's Phenomenon*** and how does this phenomenon relate to modeling with high-ordered polynomials?

Runge's Phenomenon is the phenomenon when oscillations at the edges of an interval occurs when using polynomial interpolation with polynomials of high degree.

(viii) What is the distinction between *interpolation* and *extrapolation*?

Interpolation estimates values within the range of known data, while extrapolation estimates values outside the known range.