ACTL3142 Week 4 - Logistic/Poisson Regression

Tadhg Xu-Glassop

2025T2

Recap: Linear Regression

Recall our linear regression model,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon, \qquad \epsilon | x \sim \mathcal{N}(0, \sigma^2).$$

What we are actually doing is assuming $Y|X \sim \mathcal{N}(\mu_X, \sigma^2)$, where $\mu_x = \mathbb{E}[Y|X]$, and then further assuming that μ_x is linear in X.

Recap: Linear Regression

Recall our linear regression model,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon, \qquad \epsilon | x \sim \mathcal{N}(0, \sigma^2).$$

What we are actually doing is assuming $Y|X \sim \mathcal{N}(\mu_X, \sigma^2)$, where $\mu_x = \mathbb{E}[Y|X]$, and then further assuming that μ_x is linear in X.

Note the range of the RHS is $\mathbb{R}$, which corresponds to the possible values of y on the LHS (remember, we assume $Y|X$ is normal).

Logistic Regression

Logistic Regression

Assume $Y|X \sim \mathcal{B}(p_X)$. Then, assume $\text{logit}(\mu_X)$ is linear in X, where $\mu_X = \mathbb{E}[Y|X] = p_X$. So,

$$\text{logit}(\mu_X) = \ln\left(\frac{\mathbb{P}[Y=1|X]}{1-\mathbb{P}[Y=1|X]}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

Used for binary classification.

Note $\text{logit} : (0,1) \mapsto \mathbb{R}$ and is invertible, and since the range of RHS is still $\mathbb{R}$, our model implies $\mathbb{P}[Y=1|X] \in (0,1)$, which makes sense!

Logistic Regression

Logistic Regression

Assume $Y|X \sim \mathcal{B}(p_X)$. Then, assume $\text{logit}(\mu_X)$ is linear in X, where $\mu_X = \mathbb{E}[Y|X] = p_X$. So,

$$\text{logit}(\mu_X) = \ln\left(\frac{\mathbb{P}[Y=1|X]}{1-\mathbb{P}[Y=1|X]}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

Used for binary classification.

Note $\text{logit} : (0,1) \mapsto \mathbb{R}$ and is invertible, and since the range of RHS is still $\mathbb{R}$, our model implies $\mathbb{P}[Y=1|X] \in (0,1)$, which makes sense!

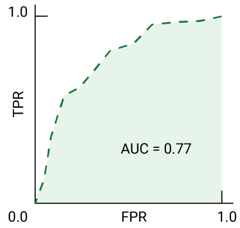
We make this a classification model with a threshold p', where we predict $Y = 1$ if $\hat{\mu}_X = \hat{p}_X > p'$.

Assessing Classification Models

Confusion Matrix - shows how many predictions we got right and what we got wrong.

	$Y = 0$	$Y = 1$	**Total**
$\hat{Y} = 0$	10	2	12
$\hat{Y} = 1$	4	14	18
Total	14	16	30

ROC Curve - plots TPR vs FPR across different thresholds. AUC is the area under this curve.



Poisson Regression

Poisson Regression

Assume $Y|X \sim \mathcal{P}(\lambda_X)$. Then, assume $\log(\mu_X)$ is linear in X, where $\mu_X = \mathbb{E}[Y|X] = \lambda_X$. So,

$$\ln(\mu_X) = \ln(\lambda_X) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

Used for count data.

Again, note $\ln : \mathbb{R}^+ \mapsto \mathbb{R}$ and is invertible, and since the range of RHS is $\mathbb{R}$ our model implies $\lambda \in \mathbb{R}^+$, which makes sense!

Remarks

Both models are modified linear regressions, so most assumptions and pitfalls carry over.

Be careful of the new interpretations of β_i's, and their effect on μ_X.

Estimation of β_i's is done via MLE using our distributional assumption.