### Generalized Linear Models

Son Nguyen

#### Generalized Linear Model

• The GLM models  $\mu=E(y)$  as follows.

$$g(\mu)=eta_0+eta_1x_1+\ldots+eta_px_p=x'eta$$

where y is assumed to follow an exponential distribution family.

- Exponential distribution family includes all the basic distribution such as normal distribution, binomial distribution, Poisson distribution...
- $g(\mu)$  is called the canonical link function
- For logistic regression, the link function is a logit function

$$g(x) = \ln\left(rac{x}{1-x}
ight)$$

### Some GLMs

$$g(\mu)=\beta_0+\beta_1x_1{+}\ldots{+}\beta_px_p=x'\beta$$

Distribution	Canonical Link Function	Mathematical Form
Normal	Identity	$g(\mu) = \mu$
Binomial	Logit	$g(\pi) = \ln[\pi/(1-\pi)]$
Poisson	Natural log	$g(\mu) = \ln \mu$
Gamma	Inverse	$g(\mu) = 1/\mu$
Inverse Gaussian	Squared inverse	$g(\mu) = 1/\mu^2$

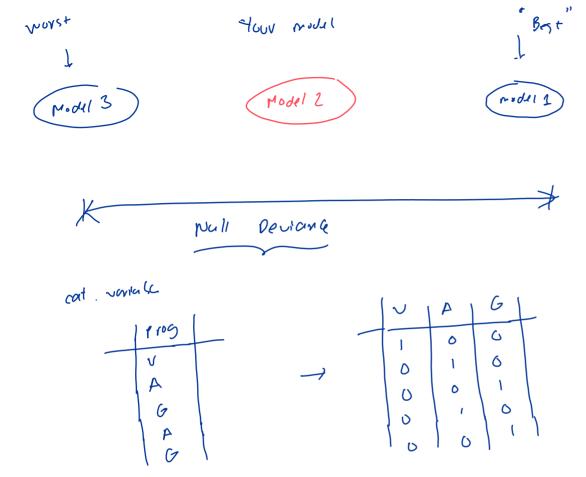
## Two Elements of GLM

- Response Assumptions
- Link Function

\_

#### Goodness of Fit: Deviance

- The deviance generalizes the Residual Sum of Squares (RSS) of the linear model
- Compare three models
- Model 1: The Perfect Model (Saturated Model)
- Model 2: Your model
- Model 3: The worst model: does not use any predictors. Also called Null Model.
- The deviance can be considered the "distance" of the model to the perfect model.
- The smaller the deviance the better the model
- Deviance 0 means the model is perfect!
- The deviance of Model 3 is also called Null Deviance.
- ullet  $R^2=1-rac{ ext{Deviance of your model}}{ ext{Deviance of the perfect model}}$



# Goodness of Fit: The Loglikelihood

- The loglikelihood of a model measures how likely the data is governed by the model.
- The higher the loglikelihood value, the better the model.

`

# Goodness of Fit: AIC

- AIC = 2k 2 \* loglikelihood of the model.
- Smaller AIC means larger loglikelihood, or better model
- AIC = Akaike information criterion