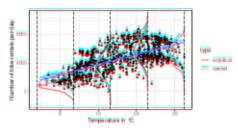
# GENERALIZED LINEAR MODEL (GLM) • Nelder and Wedderburn 1972

Problem: Target variable given feat. not always normally dist. → LM not suitable

- Target is binary (e.g., disease classification)
  - → Bernoulli / Binomial distribution
- Target is count variable (e.g., number of sold products) Poisson distribution
- Time until an event occurs (e.g., time until death) --- Gamma distribution

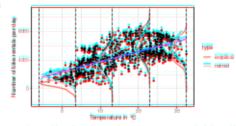




# GENERALIZED LINEAR MODEL (GLM) Nelder and Wedderburn 1972

Problem: Target variable given feat. not always normally dist. → LM not suitable

- Target is binary (e.g., disease classification)
  - → Bernoulli / Binomial distribution
- Target is count variable (e.g., number of sold products) Poisson distribution
- Time until an event occurs (e.g., time until death) Gamma distribution



Solution: GLMs - extend LMs by allowing other distributions from exponential family

$$g(\mathbb{E}(y \mid \mathbf{x})) \equiv \mathbf{x}^{\top} \theta \iff \mathbb{E}(y \mid \mathbf{x}) \equiv g^{-1}(\mathbf{x}^{\top} \theta)$$

- Link function g links linear predictor  $\mathbf{x}^{\top}\theta$  to expectation  $\mathbb{E}$  of specified Link function g links linear predictor  $\mathbf{x}^{\top}\theta$  to expectation of distribution of  $y \mid \mathbf{x}$
- $\Rightarrow$  LM is special case: Gaussian distribution for  $y \mid x$  with g as identity function
- Link function g and distribution need to be specified
- High-order and interaction effects can be manually added as in LMs
- Note: Interpretation of weights depend on link function and distribution

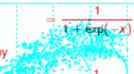


# GEM = LOGISTIC REGRESSION (GLM) \* Neitler and Wedderburn 1972

bleogistic regression ≥ GLM with Bernwalli distribution and rodit link function:

- Target is binary (e.g., disease classification)
  - → Bernoulli / Binomial distribution
- Target is cog(x)/ariloge (e.g., number of sold products
- Models probabilities for binary classification by
- Time until an event occurs (e.g., time until death)  $= E(y \mid x) = P(y = 1) = g(x \mid x)$

--- Gamma distribution



Solution: GLMs - extend LMs by allowing other distributions from exponential family

$$g(\mathbb{E}(y \mid \mathbf{x})) \neq \mathbf{x}^{\top}\theta \Leftrightarrow \mathbb{E}\mathcal{A}_{\mathbf{x}} \mid \mathbf{x}) = g^{\mathbf{T}}(\mathbf{x}^{\top}\theta)$$

- Link function g links linear predictor  $x \in \mathcal{A}$  to expectation  $\mathbb{E}$  of specified distribution of  $y \mid \mathbf{x} \mid$ 
  - → LM is special case: Gaussian distribution for y x with g as identity function.
- Link function g and distribution need to be spec score



# GEM - LOGISTIC REGRESSION (GLM) • Neider and Wedderburn 1972

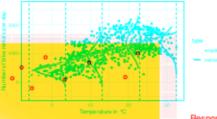
Lem: Target variable given feat, not always normally dist. LM not suitable Typically, we set the threshold to 0.5 to predict classes, e.g.,

Class 1 if  $\pi(\mathbf{x}) > 0.5$ bernoull. Binomial distribution Class 0 if  $\pi(\mathbf{x}) \leq 0.5$ 

Target is count variable

(e.g., number of sold products)

- Poisson distribution
- Tirne until an eventoccurs g., time until death) Gamma distribution



lution: GLMs - extend LMs by allowing other distributions from exponent

$$g(\mathbb{E}(y \mid \mathbf{x})) = \mathbf{x}^{\top} \theta \iff \mathbb{E}(y \mid \mathbf{x}) = g^{-1}(\mathbf{x}^{\top} \theta)$$

- $\succeq$  Link function g links linear predictor  $\mathbf{x}^{\top}\theta$  to expectation  $\mathbb{E}$  of specified distribution of  $v \mid \mathbf{x}$
- LM is special case: Gaussian distribution for  $y \mid \mathbf{x}$  with q as identity function
- ink function g and distribution need to be specified
- . High-order and interaction effects can be manually added as in LMs



## GEM - LOGISTIC REGRESSION - INTERPRETATION



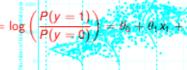
Weights  $\theta_i$  are interpreted linear as in LM (but w.r.t. log-odds)

→ difficult to comprehend

arget is count variable

(e.g., number of sold 
$$pr_{\overline{H}}(\mathbf{x})$$
cts)  
 $\rightarrow pg - odds = pg$  ( $pr_{\overline{H}}(\mathbf{x})$ )
Time until an event assure

Time until an event occurs Interpretation: leath)



Changing  $x_i$  by one unit, changes log-odds of class 1 compared to class 0 by  $\theta_i$ 

Solution: GLMs - extend LMs by allowing other distributions from exponential family

$$g(\mathbb{E}(y \mid \mathbf{x})) = \mathbf{x}^{\top} \theta \iff \mathbb{E}(y \mid \mathbf{x}) = g^{-1}(\mathbf{x}^{\top} \theta)$$

- Link function g links linear predictor  $\mathbf{x}^{\top}\theta$  to expectation  $\mathbb{E}$  of specified distribution of  $v \mid \mathbf{x}$ 
  - $\rightarrow$  LM is special case: Gaussian distribution for  $y \mid \mathbf{x}$  with g as identity function
- Link function g and distribution need to be specified
- High-order and interaction effects can be manually added as in LMs
- Note: Interpretation of weights depend on link function and distribution.



# GLM - LOGISTIC REGRESSION - INTERPRETATION

- Recall: Odds is ratio of two probabilities, odds ratio compares ratio of two odds
- Weights (b) are interpreted linear as in LM (but w.r.t. log-odds) in difficult to comprehend

Reights 
$$(\theta)$$
 are interpreted linear as in LM (but w.r.t. log-odds) difficult to comprehend 
$$\log - p d d s = \log \left( \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right) = \log \left( \frac{P(y=1)}{P(y=0)} \right) = \theta_0 + \theta_1 x_1 + \ldots + \theta_p x_p$$

Interpretation:

Changing  $x_i$  by one unit, changes log-odds of class 1 compared to class 0 by  $\theta_i$ 

- Models probabilities for binary classificati( $\mathbf{x}$ )by Odds for class 1 vs. class 0:  $odds = \frac{1}{1 \pi(\mathbf{x})} = \exp(\theta_0 + \theta_1 x_1 + \ldots + \theta_p x_p)$
- Instead of interpreting changes w.r.t. (log-odds, odds ratio is more common  $1 + \exp(-x^{-1}\theta)$

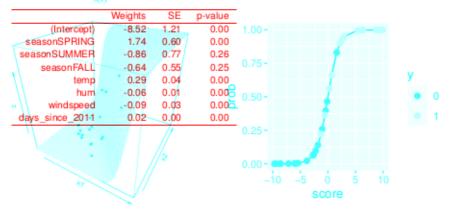
$$=\frac{\textit{odds}_{x_j+1}}{\textit{odds}}=\frac{\exp(\theta_0+\theta_1x_1+\ldots+\theta_j(x_j+1)+\ldots+\theta_px_p)}{\exp(\theta_0+\theta_1x_1+\ldots+\theta_jx_j+\ldots+\theta_px_p)}=\exp(\theta_j)$$

**Interpretation**: Changing  $x_i$  by one unit, changes the **odds ratio** for class 1 (compared to class 0) by the factor  $\exp(\theta_i)$ 



# ■GLM - LOGISTIC REGRESSION - EXAMPLE

- Creatella binary target variable for bike rental datasses, e.g.,
  - Class 1:i\*high) number of bike rentals" > 70% quantile (i.e., cnt > 5531)
  - Class 0:i\*lowto\_medium number of bike rentals" (i.e., cnt ≤ 5531)
- Fit a logistic regression model (GLM with Bernoulli distribution and logit link)





### GLM - LOGISTIC REGRESSION - EXAMPLE TATION

- Recall: Odds is quotient of two probabilities, odds ratio compares ratio of two create a binary target variable for bike rental data:
- Class 1: "high number of bike rentals" > 70% quantile (i.e., cnt > 5531)
   Weight 7: interpreted linear as in LM (but w.r.t. log didds) difficult to comprehend:

  Class 0: "low to medium number of bike rentals" (i.e., cnt ≤ 5531)
- Fit a logistic regression model (GLM with Bernoulli distribution and logit link)

	$/\pi$	<b>X</b> )	\	P(v = 1)	\
log-odds = 10	<b>N</b> eights	SE	p-value	P(	$  = \theta_0 + \theta_1 \times \cdots + \theta_p \times_p$
(Intercept)	-8.52	1.21	0.00	( [ ] = 0)	
seasonSPRING	1.74	0.60	0.00	825 G	
seasonSUMMER	-0.86	0.77	0.26	11 a	
ChseasonEALLby	on:0.64nit,	0.55.0	iges <b>0:25</b> -0	odd of class	s 1 compared to plass 0 by $\theta_I$
temp	0.29	0.04	0.00	2 F 15	
hum	-0.06	0.01	0.00	8	
windspeed	-0.09	0.03	0.00	9 8 <del></del>	
days_since_2011	0.02	0.00	0.00	-10	Temperature in "C



 If temp increases by 1°C, odds ratio for class 1 increases by factor exp(0.29) = 1.34 compared to class 0, c.p. (\hat{=} "high number of bike rentals" now 1.34 times more likely)



# GLM - LOGISTIC REGRESSION - INTERPRETATION

- Recall: Odds is quotient of two probabilities, odds ratio compares ratio of two odds
- Weights θ<sub>j</sub> interpreted linear as in LM (but w.r.t. log-odds) → difficult to comprehend

imprehend 
$$\log - odds = \log \left( \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right) = \log \left( \frac{P(y=1)}{P(y=0)} \right) = \theta_0 + \theta_1 x_1 + \ldots + \theta_p x_p$$

#### Interpretation:

Changing  $x_j$  by one unit, changes log-odds of class 1 compared to class 0 by  $\theta_j$ 

- Odds for class 1 vs. class 0:  $odds = \frac{\pi(\mathbf{x})}{1 \pi(\mathbf{x})} = \exp(\theta_0 + \theta_1 x_1 + \ldots + \theta_p x_p)$
- Instead of interpreting changes w.r.t. log-odds, it is more common to use odds ratio

$$=\frac{\textit{odds}_{x_j+1}}{\textit{odds}}=\frac{\exp(\theta_0+\theta_1x_1+\ldots+\theta_j(x_j+1)+\ldots+\theta_px_p)}{\exp(\theta_0+\theta_1x_1+\ldots+\theta_jx_j+\ldots+\theta_px_p)}=\exp(\theta_j)$$

**Interpretation**: Changing  $x_j$  by one unit, changes the **odds ratio** for class 1 (compared to class 0) by the **factor**  $\exp(\theta_j)$ 



## GLM - LOGISTIC REGRESSION - EXAMPLE

- Create a binary target variable for bike rental data:
  - Class 1: "high number of bike rentals" > 70% quantile (i.e., cnt > 5531)
  - $\bullet$  Class 0: "low to medium number of bike rentals" (i.e., cnt  $\leq$  5531)
- Fit a logistic regression model (GLM with Bernoulli distribution and logit link)

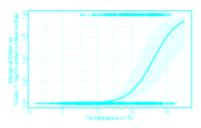
	Weights	SE	p-value
(Intercept)	-8.52	1.21	0.00
seasonSPRING	1.74	0.60	0.00
seasonSUMMER	-0.86	0.77	0.26
seasonFALL	-0.64	0.55	0.25
temp	0.29	0.04	0.00
hum	-0.06	0.01	0.00
windspeed	-0.09	0.03	0.00
days_since_2011	0.02	0.00	0.00



## GLM - LOGISTIC REGRESSION - EXAMPLE

- Create a binary target variable for bike rental data:
  - Class 1: "high number of bike rentals" > 70% quantile (i.e., cnt > 5531)
  - Class 0: "low to medium number of bike rentals" (i.e., cnt ≤ 5531)
- Fit a logistic regression model (GLM with Bernoulli distribution and logit link)

	Weights	SE	p-value
(Intercept)	-8.52	1.21	0.00
seasonSPRING	1.74	0.60	0.00
seasonSUMMER	-0.86	0.77	0.26
seasonFALL	-0.64	0.55	0.25
temp	0.29	0.04	0.00
hum	-0.06	0.01	0.00
windspeed	-0.09	0.03	0.00
days_since_2011	0.02	0.00	0.00





### Interpretation

 If temp increases by 1° C, odds ratio for class 1 increases by factor exp(0.29) = 1.34 compared to class 0, c.p. (\hat{\texi{\text{\text{\text{\text{\texitex{\text{\text{\text{\texi{\te