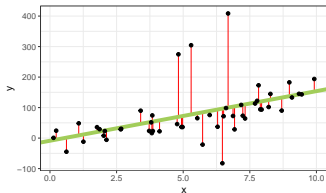


# Interpretable Machine Learning

## Extensions of Linear Regression Models



### Learning goals

- Inclusion of high-order and interaction effects
- Regularization via LASSO

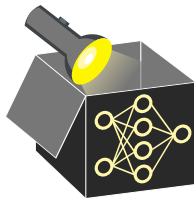
# INTERACTION AND HIGH-ORDER EFFECTS

LM Equation:  $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p + \epsilon$

Equation above can be extended (polynomial regression) by including

- **high-order effects** which have their own weights  
     $\rightsquigarrow$  e.g., quadratic effect:  $\theta_{x_j^2} \cdot x_j^2$
- **interaction effects** as the product of multiple feat.  
     $\rightsquigarrow$  e.g., 2-way interaction:  $\theta_{x_i, x_j} \cdot x_i \cdot x_j$

| Bike Data   |       |            |
|-------------|-------|------------|
| Method      | $R^2$ | adj. $R^2$ |
| Simple LM   | 0.85  | 0.84       |
| High-order  | 0.87  | 0.87       |
| Interaction | 0.96  | 0.93       |



# INTERACTION AND HIGH-ORDER EFFECTS

$$\text{LM Equation: } y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p + \epsilon$$

Equation above can be extended (polynomial regression) by including

- **high-order effects** which have their own weights

↪ e.g., quadratic effect:  $\theta_{x_j^2} \cdot x_j^2$

- **interaction effects** as the product of multiple feat.

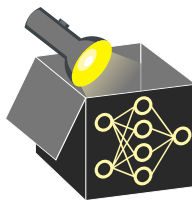
↪ e.g., 2-way interaction:  $\theta_{x_i, x_j} \cdot x_i \cdot x_j$

**Bike Data**

| Method      | $R^2$ | adj. $R^2$ |
|-------------|-------|------------|
| Simple LM   | 0.85  | 0.84       |
| High-order  | 0.87  | 0.87       |
| Interaction | 0.96  | 0.93       |

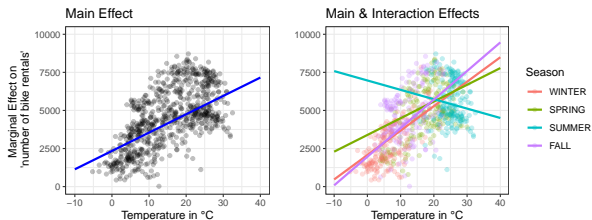
Implications of including high-order and interaction effects:

- Both make the model more flexible but also less interpretable
  - ↪ More weights to interpret
- Both need to be specified manually (inconvenient and sometimes infeasible)
  - ↪ Other ML models learn them often automatically
- Marginal effect of a feature cannot be interpreted by single weights anymore
  - ↪ Feature  $x_j$  occurs multiple times (with different weights) in equation

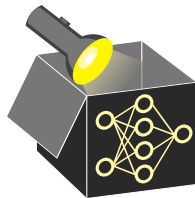


# EXAMPLE: INTERACTION EFFECT

**Example:** Interaction between temp and season will affect marginal effect of temp

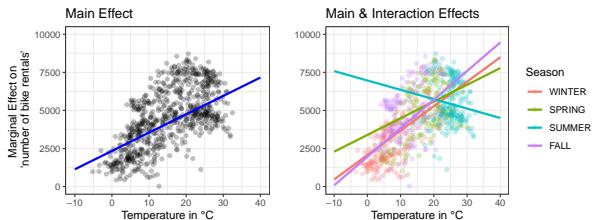


|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3453.9  |
| seasonSPRING      | 1317.0  |
| seasonSUMMER      | 4894.1  |
| seasonFALL        | -114.2  |
| temp              | 160.5   |
| hum               | -37.6   |
| windspeed         | -61.9   |
| days_since_2011   | 4.9     |
| seasonSPRING:temp | -50.7   |
| seasonSUMMER:temp | -222.0  |
| seasonFALL:temp   | 27.2    |

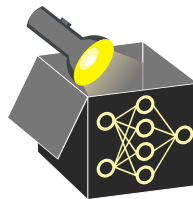


# EXAMPLE: INTERACTION EFFECT

**Example:** Interaction between temp and season will affect marginal effect of temp



|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3453.9  |
| seasonSPRING      | 1317.0  |
| seasonSUMMER      | 4894.1  |
| seasonFALL        | -114.2  |
| temp              | 160.5   |
| hum               | -37.6   |
| windspeed         | -61.9   |
| days_since_2011   | 4.9     |
| seasonSPRING:temp | -50.7   |
| seasonSUMMER:temp | -222.0  |
| seasonFALL:temp   | 27.2    |

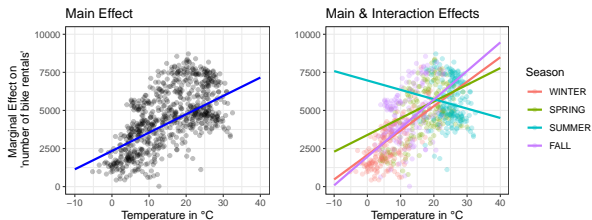


**Interpretation:** If temp increases by 1 °C, bike rentals

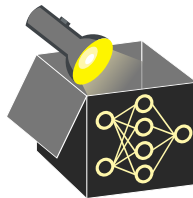
- increase by 160.5 in WINTER (reference)

# EXAMPLE: INTERACTION EFFECT

**Example:** Interaction between temp and season will affect marginal effect of temp



|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3453.9  |
| seasonSPRING      | 1317.0  |
| seasonSUMMER      | 4894.1  |
| seasonFALL        | -114.2  |
| temp              | 160.5   |
| hum               | -37.6   |
| windspeed         | -61.9   |
| days_since_2011   | 4.9     |
| seasonSPRING:temp | -50.7   |
| seasonSUMMER:temp | -222.0  |
| seasonFALL:temp   | 27.2    |

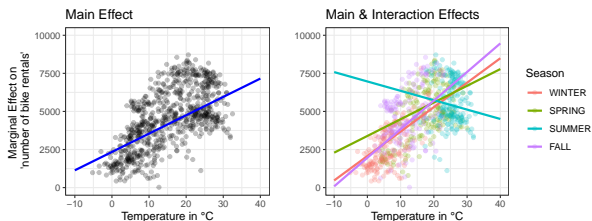


**Interpretation:** If temp increases by 1 °C, bike rentals

- increase by 160.5 in WINTER (reference)
- increase by 109.8 (= 160.5 - 50.7) in SPRING

# EXAMPLE: INTERACTION EFFECT

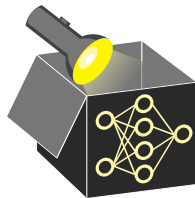
**Example:** Interaction between temp and season will affect marginal effect of temp



|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3453.9  |
| seasonSPRING      | 1317.0  |
| seasonSUMMER      | 4894.1  |
| seasonFALL        | -114.2  |
| temp              | 160.5   |
| hum               | -37.6   |
| windspeed         | -61.9   |
| days_since_2011   | 4.9     |
| seasonSPRING:temp | -50.7   |
| seasonSUMMER:temp | -222.0  |
| seasonFALL:temp   | 27.2    |

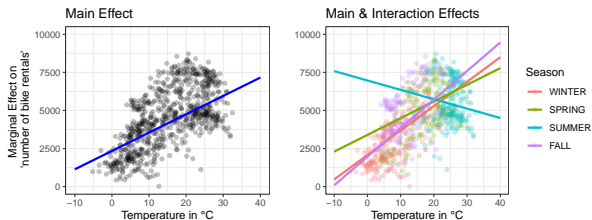
**Interpretation:** If temp increases by 1 °C, bike rentals

- increase by 160.5 in WINTER (reference)
- increase by 109.8 (= 160.5 - 50.7) in SPRING
- decrease by -61.5 (= 160.5 - 222) in SUMMER

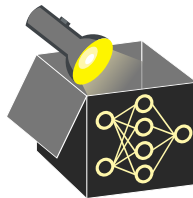


# EXAMPLE: INTERACTION EFFECT

**Example:** Interaction between temp and season will affect marginal effect of temp



|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3453.9  |
| seasonSPRING      | 1317.0  |
| seasonSUMMER      | 4894.1  |
| seasonFALL        | -114.2  |
| temp              | 160.5   |
| hum               | -37.6   |
| windspeed         | -61.9   |
| days_since_2011   | 4.9     |
| seasonSPRING:temp | -50.7   |
| seasonSUMMER:temp | -222.0  |
| seasonFALL:temp   | 27.2    |



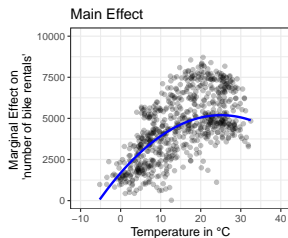
**Interpretation:** If temp increases by 1 °C, bike rentals

- increase by 160.5 in WINTER (reference)
- increase by 109.8 (= 160.5 - 50.7) in SPRING
- decrease by -61.5 (= 160.5 - 222) in SUMMER
- increase by 187.7 (= 160.5 + 27.2) in FALL



# EXAMPLE: QUADRATIC EFFECT

**Example:** Adding quadratic effect for temp

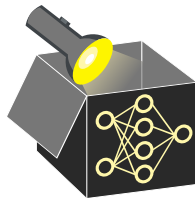


**Interpretation:** Not linear anymore!

- temp depends on two weights:

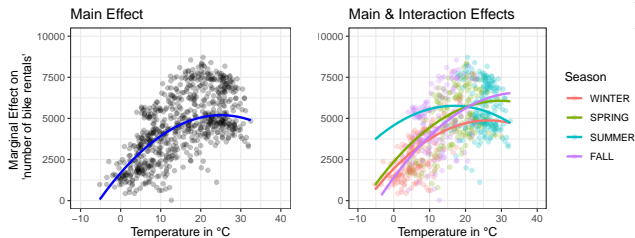
$$280.2 \cdot x_{temp} - 5.6 \cdot x_{temp}^2$$

|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3094.1  |
| seasonSPRING      | 619.2   |
| seasonSUMMER      | 284.6   |
| seasonFALL        | 123.1   |
| hum               | -36.4   |
| windspeed         | -65.7   |
| days_since_2011   | 4.7     |
| temp              | 280.2   |
| temp <sup>2</sup> | -5.6    |



# EXAMPLE: QUADRATIC EFFECT

**Example:** Adding quadratic effect for temp (left) and interaction with season (right)



|                                | Weights |
|--------------------------------|---------|
| (Intercept)                    | 3802.1  |
| seasonSPRING                   | -1345.1 |
| seasonSUMMER                   | -6006.3 |
| seasonFALL                     | -681.4  |
| hum                            | -38.9   |
| windspeed                      | -64.1   |
| days_since_2011                | 4.8     |
| temp                           | 39.1    |
| temp <sup>2</sup>              | 8.6     |
| seasonSPRING:temp              | 407.4   |
| seasonSPRING:temp <sup>2</sup> | -18.7   |
| seasonSUMMER:temp              | 801.1   |
| seasonSUMMER:temp <sup>2</sup> | -27.2   |
| seasonFALL:temp                | 217.4   |
| seasonFALL:temp <sup>2</sup>   | -11.3   |

**Interpretation:** Not linear anymore!

- temp depends on multiple weights due to season:

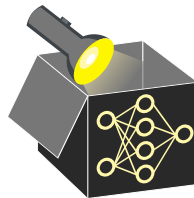
↪ WINTER:  $39.1 \cdot x_{temp} + 8.6 \cdot x_{temp}^2$

↪ SPRING:  $(39.1+407.4) \cdot x_{temp} + (8.6-18.7) \cdot x_{temp}^2$

↪ SUMMER:

$(39.1+801.1) \cdot x_{temp} + (8.6-27.2) \cdot x_{temp}^2$

↪ FALL:  $(39.1+217.4) \cdot x_{temp} + (8.6-11.3) \cdot x_{temp}^2$

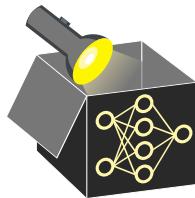
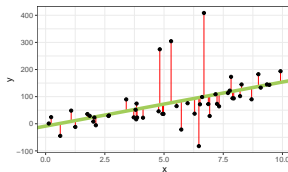
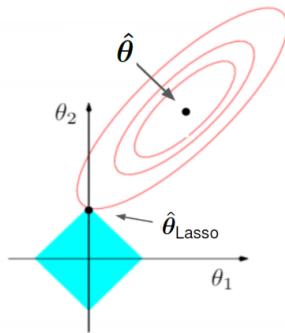


# REGULARIZATION VIA LASSO

► Tibshirani (1996)

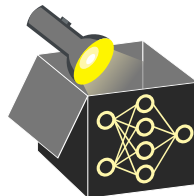
- LASSO adds an  $L_1$ -norm penalization term ( $\lambda ||\theta||_1$ ) to least squares optimization problem
  - ↪ Shrinks some feature weights to zero (feature selection)
  - ↪ Sparser models (fewer features): more interpretable
- Penalization parameter  $\lambda$  must be chosen (e.g., by CV)

$$\min_{\theta} \left( \underbrace{\frac{1}{n} \sum_{i=1}^n (y^{(i)} - \mathbf{x}^{(i)\top} \theta)^2}_{\text{Least square estimate for LM}} + \lambda ||\theta||_1 \right)$$



# REGULARIZATION VIA LASSO

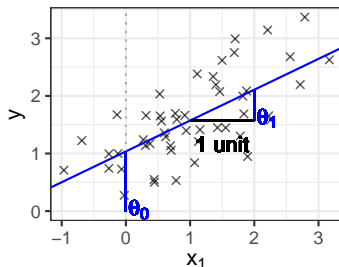
► Tibshirani (1996)



**Example** (interpretation of weights analogous to LM):

- LASSO with main effects and interaction temp with season
- $\lambda$  is chosen  $\rightsquigarrow$  6 selected features ( $\neq 0$ )
- LASSO shrinks weights of single categories separately (due to dummy encoding)
  - $\rightsquigarrow$  No feature selection of whole categorical features (only w.r.t. category levels)
  - $\rightsquigarrow$  Solution: group LASSO

► Yuan and Lin (2006)



|                   | Weights |
|-------------------|---------|
| (Intercept)       | 3135.2  |
| seasonSPRING      | 767.4   |
| seasonSUMMER      | 0.0     |
| seasonFALL        | 0.0     |
| temp              | 116.7   |
| hum               | -28.9   |
| windspeed         | -50.5   |
| days_since_2011   | 4.8     |
| seasonSPRING:temp | 0.0     |
| seasonSUMMER:temp | 0.0     |
| seasonFALL:temp   | 30.2    |