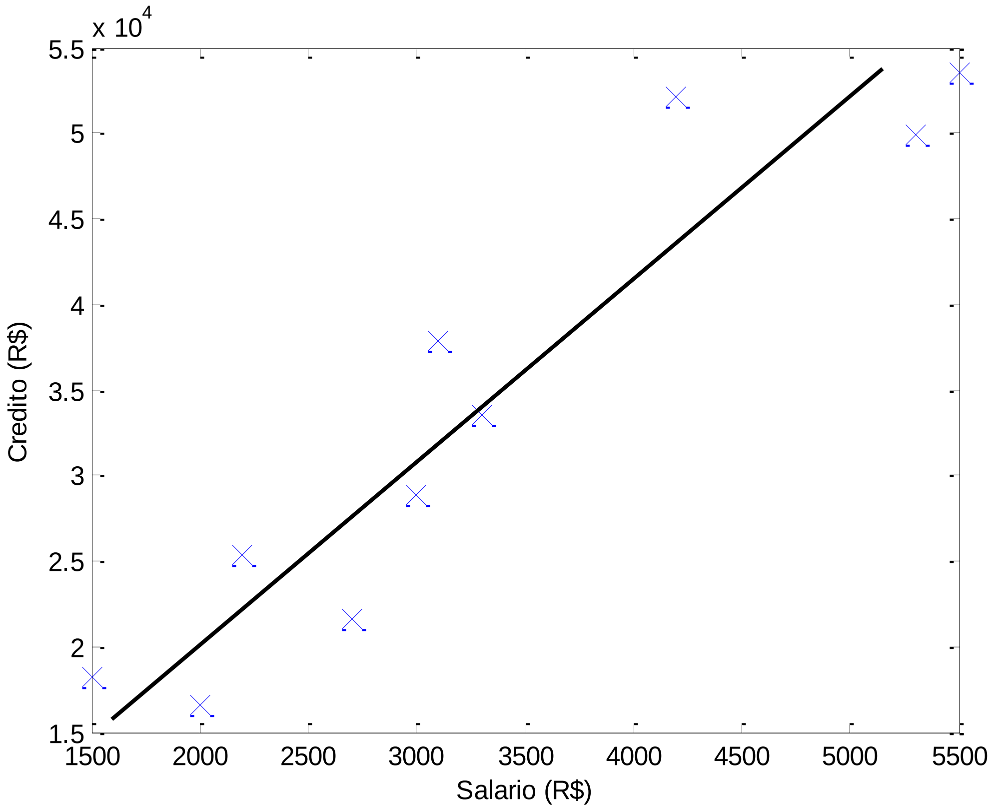
Aprendizagem Automática

João Paulo Pordeus Gomes

Aula Anterior

Regressão Linear

- $\bar{y}_i = w_1 x_i + w_0$



Regressão Linear Multivariada

- Em diversos problemas é necessário utilizar mais de uma variável x para tentar explicar a variável de saída y
  - Exemplo
    - Concessão de crédito

Salário (R$)	Dívida (R$)	Crédito (R$)
1500	0	16500
2000	4000	12000
3000	2000	28000
...		...
5500	1700	52000

Gradiente Descendente Estocástico

- Regressão Linear Multivariada
  - $\overline{y_i} = \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{i}}$
- Regra de Aprendizado
  - $\mathbf{w} = \mathbf{w} + \alpha e_i \boldsymbol{x}_{\boldsymbol{i}}$

Regressão Linear (em lote - *batch*)

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de Aprendizado
  - $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$

Gradiente Descendente Estocástico

- Define α pequeno
- Utiliza todo o conjunto de dados e atualiza os pesos
  - $w_0 = w_0 + \alpha e_i$
  - $w_1 = w_1 + \alpha e_i x_i$
- Faz permutação nos dados
- Repete o procedimento diversas vezes (épocas)

Criando Modelos Não Lineares

Modelos não-Lineares com Regressão Linear

- É possível criar regressões não lineares através da formulação linear do problema.
  - Relação entre as variáveis é reconhecidamente não linear (quadrática, cúbica ...)

Modelos não-Lineares com Regressão Linear

- É possível criar regressões não lineares através da formulação linear do problema
- Criar novos atributos

Modelos não-Lineares com Regressão Linear

- É possível criar regressões não lineares através da formulação linear do problema
- Criar novos atributos

Salário (R$)	Crédito (R$)
1500	16500
2000	18000
3000	28000
...	...
5500	52000

Modelos não-Lineares com Regressão Linear

- É possível criar regressões não lineares através da formulação linear do problema
- Criar novos atributos

Salário (R$)	Salário^2 (R$^{2})	Crédito (R$)
1500	1500^2	16500
2000	2000^2	18000
3000	3000^2	28000
...		...
5500	5500^2	52000

Modelos não-Lineares com Regressão Linear

- É possível criar regressões não lineares através da formulação linear do problema
- Criar novos atributos
  - $\overline{y_i} = \boldsymbol{w}^T \boldsymbol{x}_i$
    - $\boldsymbol{x_1} = [1\ \ 1500\ 1500^2\ ]T$, $\boldsymbol{x_2} = [1\ \ 2000\ \ 2000^2]^T$, …
    - $\boldsymbol{w}^T = [w_0\ \ w_1\ \ w_2]$

Salário (R$)	Salário² (R$²)	Crédito (R$)
1500	1500^2	16500
2000	2000^2	18000
3000	3000^2	28000
...		...
5500	5500^2	52000

Regularização

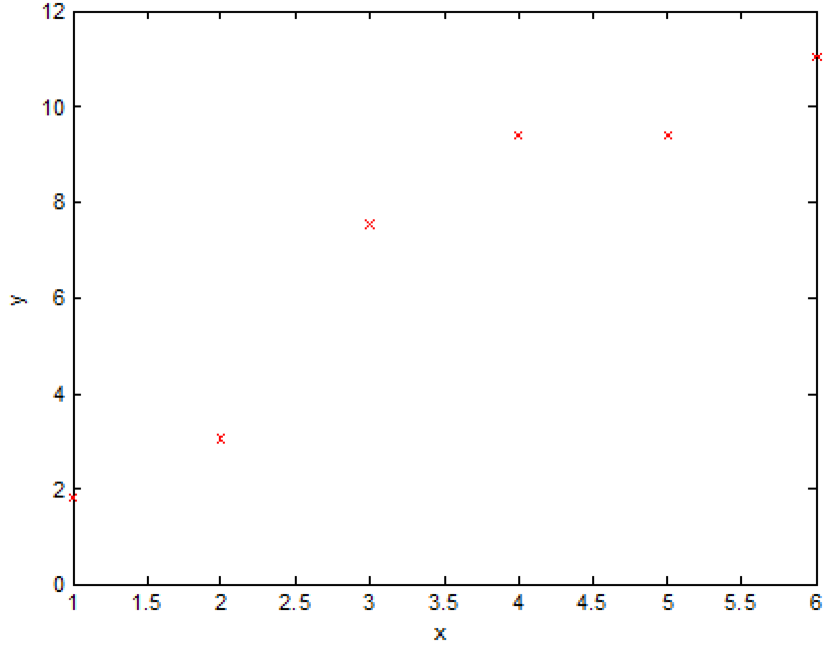
Regressão Linear

Overfitting

- Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros

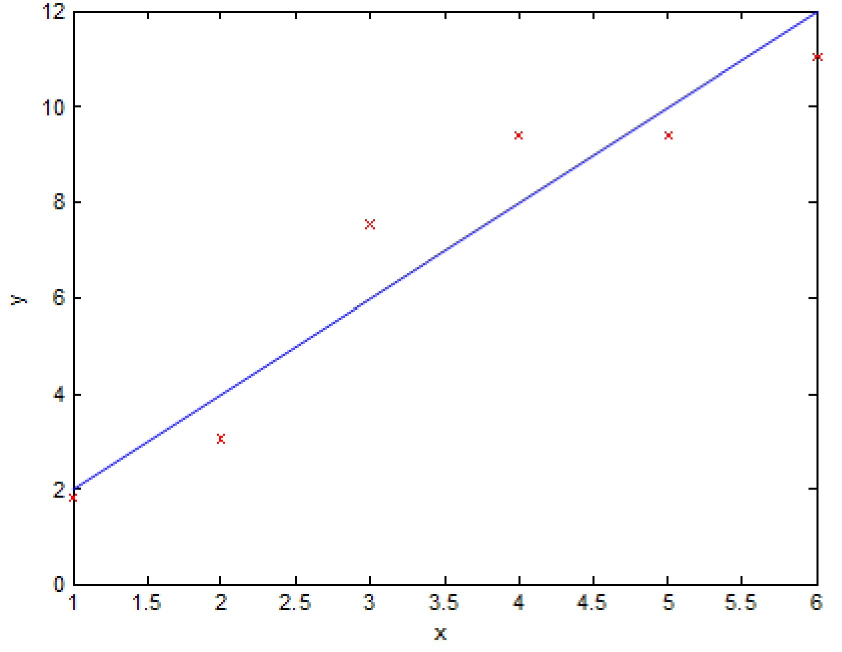
Overfitting

- Exemplo



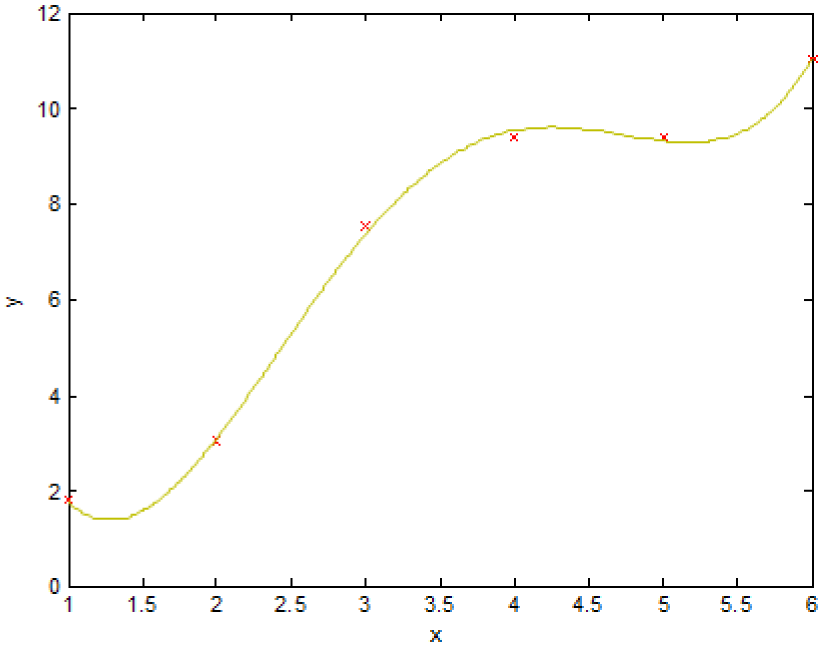
Overfitting

- Exemplo
  - Erro = 6.54



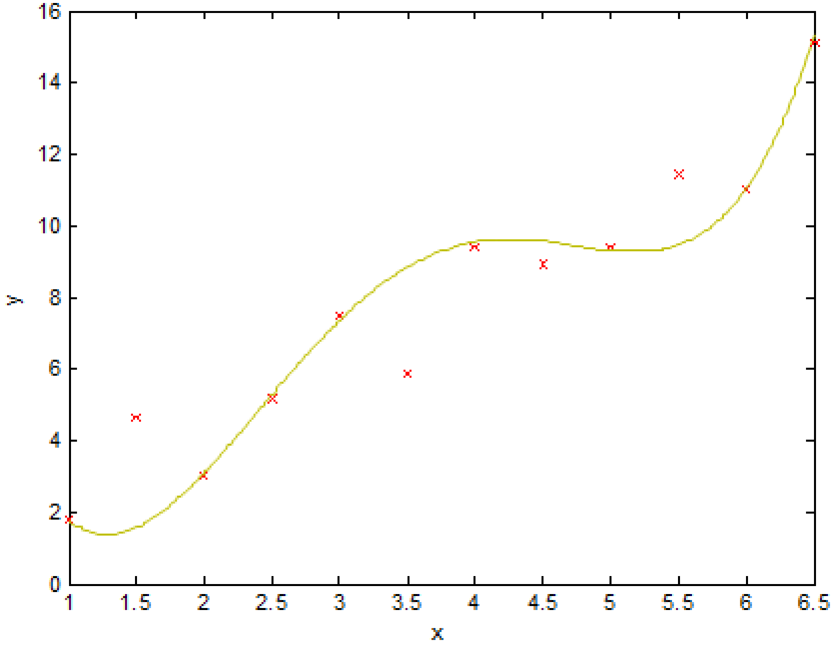
Overfitting

- Exemplo
  - Erro = 2.24



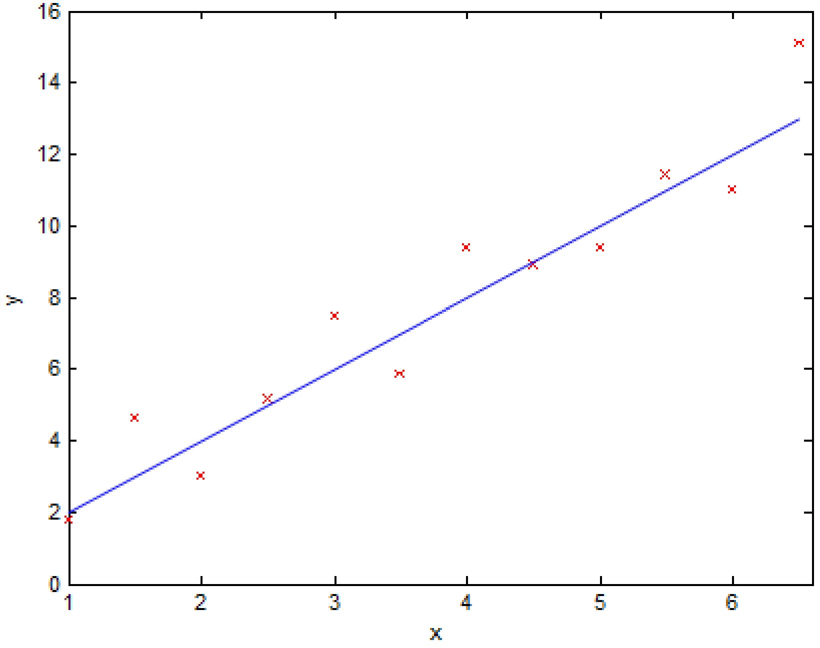
Overfitting

- Exemplo
  - Erro = 27.77



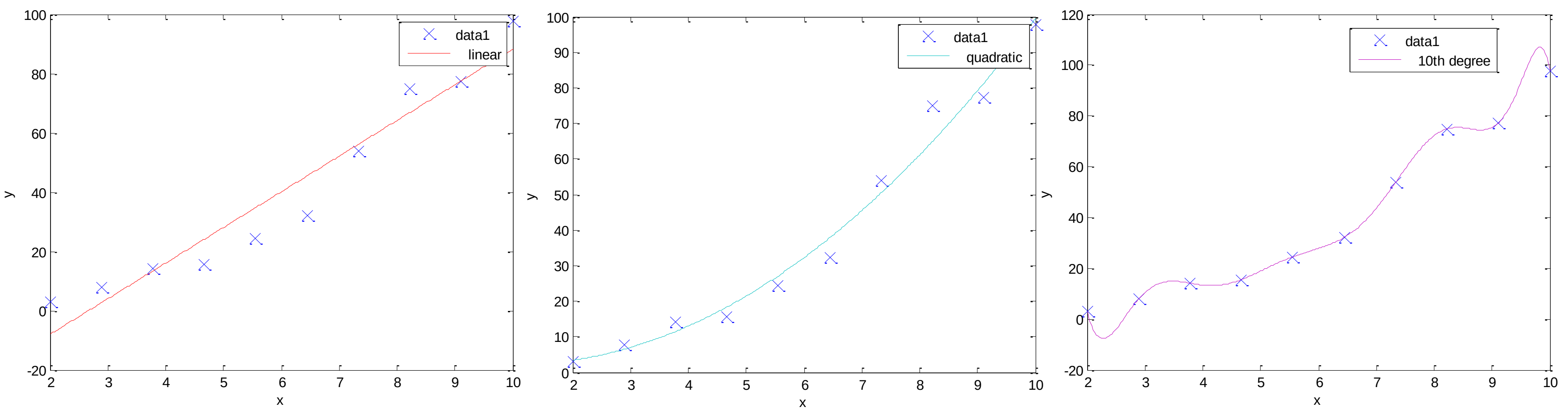
Overfitting

- Exemplo
  - Erro = 15.18



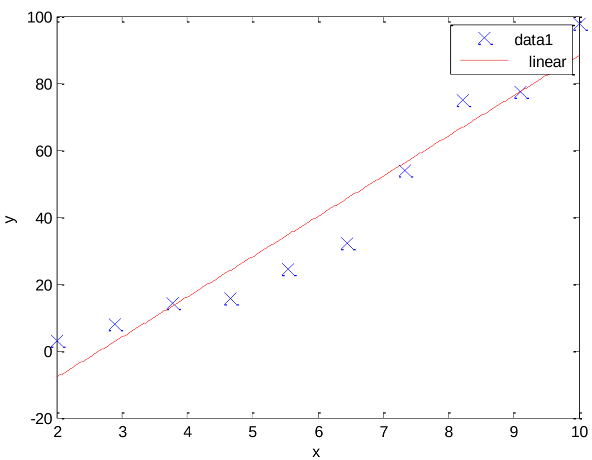
Overfitting

- Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros

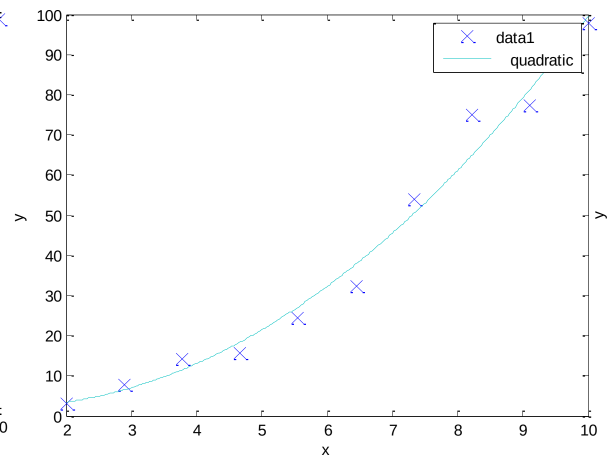


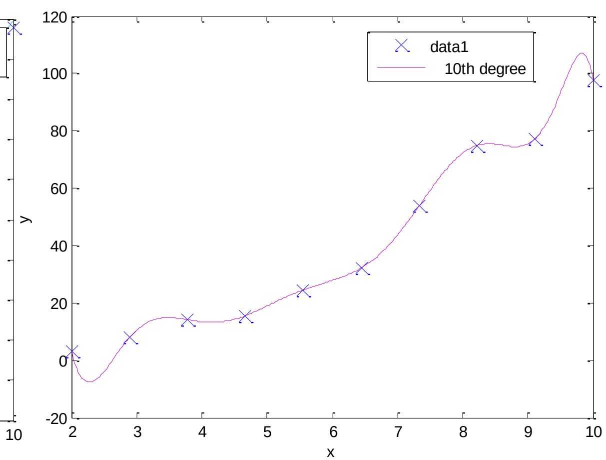
Overfitting

- Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros



Alto Bias





Alto Variância

Evitar Overfitting

Evitar Overfitting

- Conjuntos de treino e teste

Evitar Overfitting

- Conjuntos de treino e teste
- Regularização

Tamanho do Modelo

- Como ajustar o tamanho do modelo ?
  - Exemplo
    - Diversas Variáveis
      - Concessão de crédito com muitas variáveis de entrada
      - $\bar{y} = w_1x_1 + w_0$
      - $\bar{y} = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_0$
    - Modelo de grau maior
      - $\bar{y} = w_1x_1 + w_0$
      - $\bar{y} = w_1x_1 + w_1{x_1}^2 + w_1{x_1}^3 + w_1{x_1}^4 + w_0$

Regularização

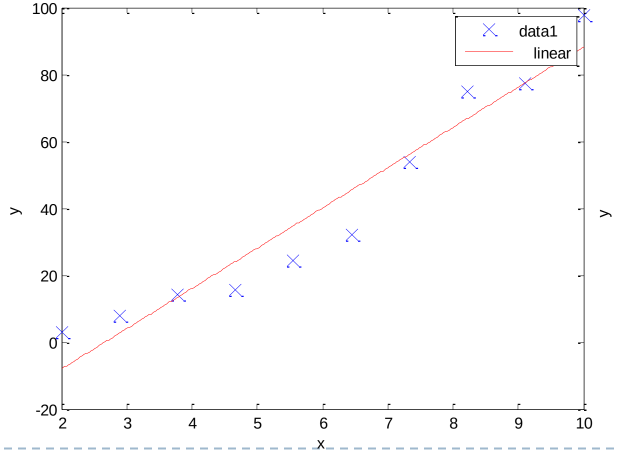
- Diminuir o valor dos coeficientes do modelo associados a variáveis que menos influenciam no resultado.

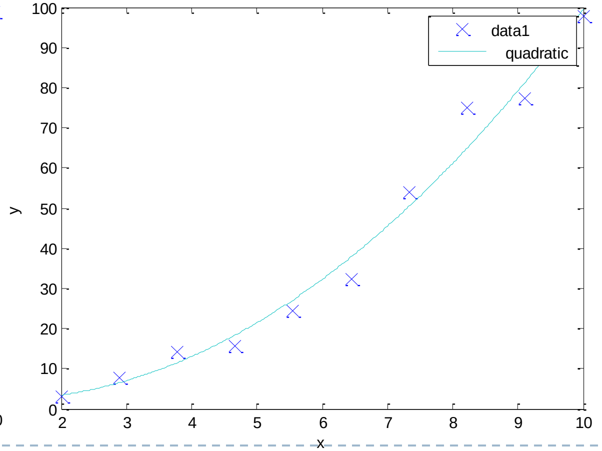
Regularização

- Diminuir o valor dos coeficientes do modelo associados a variáveis que menos influenciam no resultado.
- Modelo com muitos parâmetros ficará semelhante a um modelo com menos parâmetros
  - $\bar{y} = w_1 x_1 + w_0$
  - $\bar{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_0$

Regularização

- Diminuir o valor dos coeficientes do modelo associados a variáveis que menos influenciam no resultado.
- Modelo com muitos parâmetros ficará semelhante a um modelo com menos parâmetros
  - $\bar{y} = w_1 x_1 + w_0$
  - $\bar{y} = w_1 x_1 + w_1 {x_1}^2 + w_0$





Regularização

- Regressão Linear
  - Função Objetivo
    - $J(\boldsymbol{w}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \overline{y_i})^2$

Regularização

- Regressão Linear
  - Função Objetivo
    - $J(\boldsymbol{w}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \overline{y_i})^2$
- Regressão Linear com regularização
  - Função Objetivo
    - $J(\boldsymbol{w}) = \frac{1}{2n}[\sum_{i=1}^{n}(y_i - \overline{y_i})^2 + \lambda\sum_{j=1}^{m}{w_j}^2]$

Regularização

- $min_{\mathbf{w}} J(\boldsymbol{w}) = \frac{1}{2n} [\sum_{i=1}^{n} (y_i - \overline{y_i})^2 + \lambda \sum_{j=1}^{m} w_j{}^2]$

Regularização

- $min_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{1}{2n}[\sum_{i=1}^{n}(y_i - \overline{y_i})^2 + \lambda \sum_{j=1}^{m} w_j{}^2]$
- Utilizando o método dos mínimos quadrados
  - $\frac{\partial J}{\partial w_0} = \frac{1}{2n} 2[\sum_{i=1}^{n} e_i(-1)] = -\frac{1}{n}\sum_{i=1}^{n} e_i$

Regularização

- $min_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{1}{2n}[\sum_{i=1}^{n}(y_i - \overline{y_i})^2 + \lambda \sum_{j=1}^{m} w_j{}^2]$
- Utilizando o método dos mínimos quadrados
  - $\frac{\partial J}{\partial w_0} = \frac{1}{2n} 2[\sum_{i=1}^{n} e_i(-1)] = -\frac{1}{n}\sum_{i=1}^{n} e_i$
  - $\frac{\partial J}{\partial w_j} = \frac{1}{2n} 2[\sum_{i=1}^{n} e_i\left(-x_{ij}\right) + \frac{\lambda}{n} w_j] = -\frac{1}{n}\sum_{i=1}^{n} e_i x_{ij} - \frac{\lambda}{n} w_j$

Regularização

- $\frac{\partial J}{\partial w_0} = -\frac{1}{n}\sum_{i=1}^{n} e_i$
- $\frac{\partial J}{\partial w_j} = -\frac{1}{n}\sum_{i=1}^{n} e_i x_{ij} + \frac{\lambda}{n} w_j$

- As equações de atualização dos pesos serão:
  - $w_0 = w_0 + \alpha \frac{1}{n}\sum_{i=1}^{n} e_i$
  - $w_j = w_j + \alpha[\frac{1}{n}\sum_{i=1}^{n} e_i\, x_{ij} - \frac{\lambda}{n} w_j]$

Regularização

- Efeito do λ
  - Função Objetivo
    - $J(\boldsymbol{w}) = \frac{1}{2n}[\sum_{i=1}^{n}(y_i - \overline{y_i})^2 + \lambda \sum_{j=1}^{m} w_j{}^2]$
  - Atualização
    - $w_0 = w_0 + \alpha \frac{1}{n} \sum_{i=1}^{n} e_i$
    - $w_j = w_j + \alpha[\frac{1}{n} \sum_{i=1}^{n} e_i \, x_{ij} - \frac{\lambda}{n} w_j]$

Regressão Linear

- Gradiente Descendente
- Mínimos Quadrados (batch)

Mínimos Quadrados

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{Xw}$
- Função de custo
  - $J(\boldsymbol{w}\ ) = \frac{1}{2}(Y - \overline{\boldsymbol{Y}})^{\boldsymbol{T}}(Y - \overline{\boldsymbol{Y}})$
  - $J(\boldsymbol{w}\ ) = \frac{1}{2}(Y - \boldsymbol{Xw})^{\boldsymbol{T}}(Y - \boldsymbol{Xw})$

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Função de custo
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \overline{\boldsymbol{Y}})^{\boldsymbol{T}}(Y - \overline{\boldsymbol{Y}}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Função de custo
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \overline{\boldsymbol{Y}})^{\boldsymbol{T}}(Y - \overline{\boldsymbol{Y}}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
- Derivando em relação a **w**

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Função de custo
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \overline{\boldsymbol{Y}})^{\boldsymbol{T}}(Y - \overline{\boldsymbol{Y}}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
- Derivando em relação a **w**
  - $\frac{1}{2}2[(-\boldsymbol{X})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}] = \boldsymbol{0}$

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Função de custo
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \overline{\boldsymbol{Y}})^{\boldsymbol{T}}(Y - \overline{\boldsymbol{Y}}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
  - $J(\boldsymbol{w}) = \frac{1}{2}[(Y - \boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}]$
- Derivando em relação a **w**
  - $\frac{1}{2}2[(-\boldsymbol{X})^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w}] = \boldsymbol{0}$
  - $-\boldsymbol{X}^{\boldsymbol{T}}(Y - \boldsymbol{X}\boldsymbol{w}) + \lambda \boldsymbol{w} = \boldsymbol{0}$
  - $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{w} + \lambda \boldsymbol{w} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
  - $(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)\boldsymbol{w} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
  - $\boldsymbol{w} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de ajuste dos pesos
  - $\boldsymbol{w} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de ajuste dos pesos
  - $\boldsymbol{w} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
- Não utilizar regularização no termo w_0

Mínimos Quadrados Regularizado

- Modelo
  - $\overline{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de ajuste dos pesos
  - $\boldsymbol{w} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
- Não utilizar regularização no termo w_0
  - Fazer o primeiro termo de λI igual a zero

Mínimos Quadrados Regularizado

- Modelo
  - $\bar{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de ajuste dos pesos
  - $\boldsymbol{w} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
- Não utilizar regularização no termo w_0
  - Fazer o primeiro termo de λI igual a zero
  - Exemplo
    - $\lambda I = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$

Mínimos Quadrados Regularizado

- Modelo
  - $\bar{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{w}$
- Regra de ajuste dos pesos
  - $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
- Não utilizar regularização no termo w_0
  - Fazer o primeiro termo de λI igual a zero
  - Exemplo
    - $\lambda I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$

Dúvidas ?