APRENDIZADO DE MÁQUINA

FUNÇÃO DE PERDA (I)

Zenilton K. G. Patrocínio Jr

zenilton@pucminas.br

Função de Perda

Pode não haver um valor alvo y "verdadeiro" para uma observação x, ou seja, pode haver vários "diferentes" valores de y para o mesmo x

Função de Perda

Pode não haver um valor alvo y “verdadeiro” para uma observação x, ou seja, pode haver vários “diferentes” valores de y para o mesmo x

Também pode haver ruído ou efeitos não modelados no conjunto de dados; portanto, mesmo se houver um único y para um dado x, pode ser impossível predizê-lo com exatidão

Função de Perda

Pode não haver um valor alvo y “verdadeiro” para uma observação x, ou seja, pode haver vários “diferentes” valores de y para o mesmo x

Também pode haver ruído ou efeitos não modelados no conjunto de dados; portanto, mesmo se houver um único y para um dado x, pode ser impossível predizê-lo com exatidão

Em vez disso, tenta-se predizer um valor “próximo” do valor alvo observado

Função de Perda

Neste caso, usa-se durante o treinamento uma **função de perda** para medir a proximidade de tais predições feitas em relação aquelas do conjunto de dados que foram previamente observadas

Função de Perda

Neste caso, usa-se durante o treinamento uma **função de perda** para medir a proximidade de tais predições feitas em relação aquelas do conjunto de dados que foram previamente observadas

Uma **função de perda** mede a diferença entre uma predição do valor alvo e o valor disponível no conjunto de treinamento

Função de Perda

Neste caso, usa-se durante o treinamento uma **função de perda** para medir a proximidade de tais predições feitas em relação aquelas do conjunto de dados que foram previamente observadas

Uma **função de perda** mede a diferença entre uma predição do valor alvo e o valor disponível no conjunto de treinamento

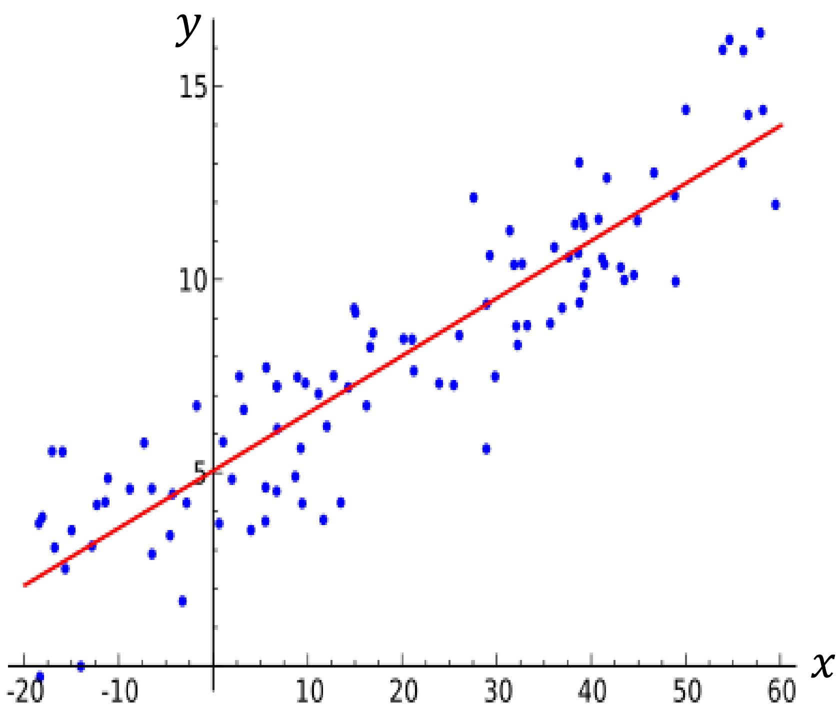
Exemplo: Perda quadrática (ou "***squared loss***") $L_2(\hat{y}, y) = (\hat{y} - y)^2$ em que $\hat{y} = f(x)$ representa a predição feita pelo modelo para o par de dados (x, y)

Exemplo – Regressão Linear

Nesse tipo de modelo bem simples, $\hat{y} = f(x) = ax + b$ sendo x um valor real, enquanto a e b são constantes reais

Exemplo – Regressão Linear

Nesse tipo de modelo bem simples, $\hat{y} = f(x) = ax + b$ sendo x um valor real, enquanto a e b são constantes reais



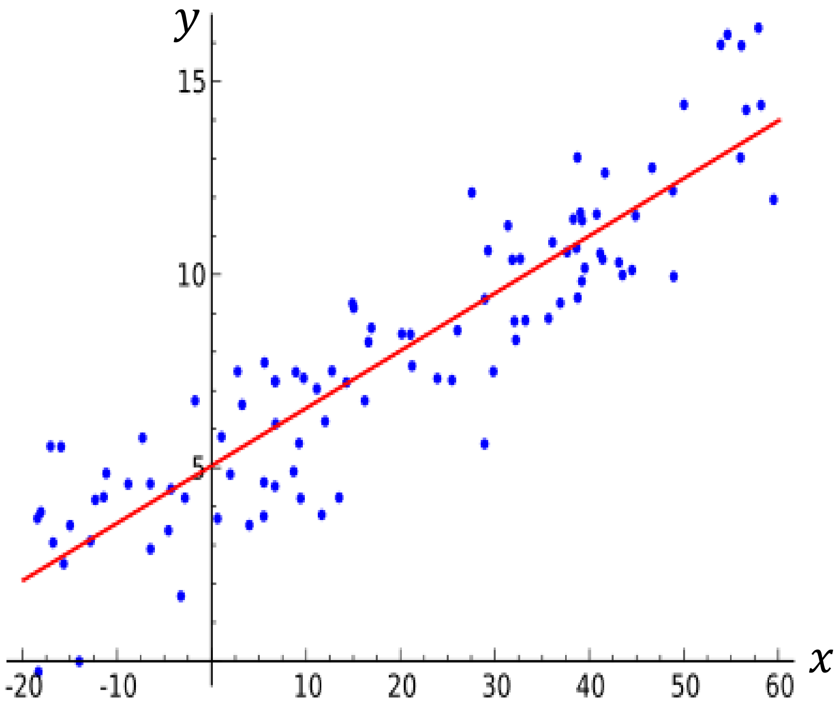
Os n pares de dados (x, y) são os pontos azuis

Já o modelo é representado pela linha vermelha

Exemplo – Regressão Linear

Nesse tipo de modelo bem simples, $\hat{y} = f(x) = ax + b$ sendo x um valor real, enquanto a e b são constantes reais

A perda quadrática $L_2(\hat{y}, y) = (\hat{y} - y)^2$ avalia a diferença (ou distância) entre predição $\hat{y}$ e dado y



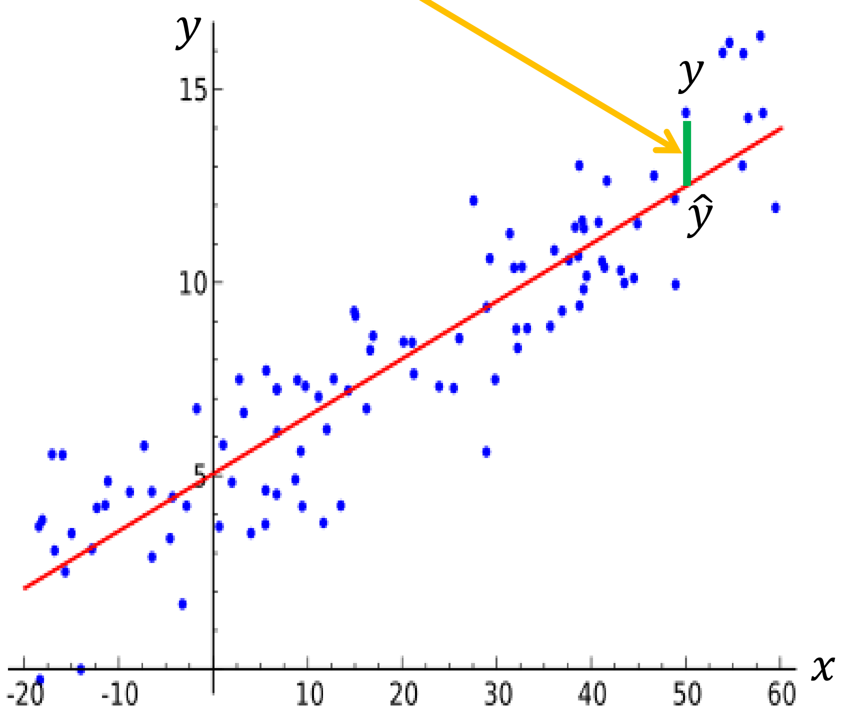
Os n pares de dados (x, y) são os pontos azuis

Já o modelo é representado pela linha vermelha

Exemplo – Regressão Linear

Nesse tipo de modelo bem simples, $\hat{y} = f(x) = ax + b$ sendo x um valor real, enquanto a e b são constantes reais

A perda quadrática $L_2(\hat{y}, y) = (\hat{y} - y)^2$ avalia a diferença (ou distância) entre predição $\hat{y}$ e dado y



Os n pares de dados (x, y) são os pontos azuis

Já o modelo é representado pela linha vermelha

Exemplo – Regressão Linear

A perda total pode ser obtida somando-se a perda em todos os n pares

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Exemplo – Regressão Linear

A perda total pode ser obtida somando-se a perda em todos os n pares

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Uma vez que a predição $\hat{y}_i$ para um x_i qualquer é dada por $ax_i + b$, pode-se rescrever a perda total como

$$L = \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

Exemplo – Regressão Linear

A perda total pode ser obtida somando-se a perda em todos os n pares

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Uma vez que a predição $\hat{y}_i$ para um x_i qualquer é dada por $ax_i + b$, pode-se rescrever a perda total como

$$L = \sum_{i=1}^{n} (ax_i + b - y_i)^2$$

Para se encontrar os valores ótimos de a e b, pode-se aplicar o cálculo diferencial, igualando à zero as derivadas da perda total em relação a e b, isto é

$$\frac{dL}{da} = 0 \quad \text{e} \quad \frac{dL}{db} = 0$$

Exemplo – Regressão Linear

Dessa forma, para se minimizar a perda e obter os valores ótimos de a e b, deve-se fazer

$$\frac{dL}{da} = 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i = 0$$

$$\frac{dL}{db} = 2a\sum_{i=1}^{n} x_i + 2bn - 2\sum_{i=1}^{n} y_i = 0$$

Exemplo – Regressão Linear

Dessa forma, para se minimizar a perda e obter os valores ótimos de a e b, deve-se fazer

$$\frac{dL}{da} = 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i = 0$$

$$\frac{dL}{db} = 2a\sum_{i=1}^{n} x_i + 2bn - 2\sum_{i=1}^{n} y_i = 0$$

Como os somatórios nessas equações são constantes para os n pares de dados, tem-se duas equações **lineares** em a e b, que podem ser facilmente resolvidas

Exemplo – Regressão Linear

Dessa forma, para se minimizar a perda e obter os valores ótimos de a e b, deve-se fazer

$$\frac{dL}{da} = 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i = 0$$

$$\frac{dL}{db} = 2a\sum_{i=1}^{n} x_i + 2bn - 2\sum_{i=1}^{n} y_i = 0$$

Como os somatórios nessas equações são constantes para os n pares de dados, tem-se duas equações **lineares** em a e b, que podem ser facilmente resolvidas

O modelo $\boldsymbol{f(x) = ax + b}$ com estes valores é o **modelo de perda mínima**

Minimização de Risco

Recapitulando... obteve-se o valor das constantes a e b que minimizam a perda quadrática sobre alguns dados que ***nós já possuíamos***

Minimização de Risco

Recapitulando... obteve-se o valor das constantes a e b que minimizam a perda quadrática sobre alguns dados que ***nós já possuíamos***

Porém o que se quer na verdade é predizer os valores de y para observações x que ***nós ainda não dispomos***, isto é, gostaríamos de minimizar a perda esperada sobre novos dados, ou ainda, $\mathbb{E}[(\hat{y} - y)^2]$

Minimização de Risco

Recapitulando... obteve-se o valor das constantes a e b que minimizam a perda quadrática sobre alguns dados que ***nós já possuíamos***

Porém o que se quer na verdade é predizer os valores de y para observações x que ***nós ainda não dispomos***, isto é, gostaríamos de minimizar a perda esperada sobre novos dados, ou ainda, $\mathbb{E}[(\hat{y} - y)^2]$

Tal perda esperada é denominada **risco**

Minimização de Risco

Na verdade, minimizou-se um valor de perda obtido sobre um número finito de dados disponíveis que é chamado de **risco empírico**

Minimização de Risco

Na verdade, minimizou-se um valor de perda obtido sobre um número finito de dados disponíveis que é chamado de **risco empírico**

Dessa forma, o aprendizado de máquina aproxima modelos que minimizam o risco por meio de modelos que minimizem o risco empírico

Minimização de Risco

Na verdade, minimizou-se um valor de perda obtido sobre um número finito de dados disponíveis que é chamado de **risco empírico**

Dessa forma, o aprendizado de máquina aproxima modelos que minimizam o risco por meio de modelos que minimizem o risco empírico

Geralmente, minimizar o risco empírico (perda de dados) em vez do risco real funciona bem, mas pode falhar se:

- A amostra de dados for enviesada, ou
- Não houver dados suficientes para estimar com precisão os parâmetros do modelo