Linear regression

Outline

Linear Regression

- Linear model
- Mean squared error function
- Parameter estimation.
- Gradient methods.
- On-line regression techniques.
- Linear additive models
- Statistical model of linear regression
- Regularized linear regression
Supervised learning

Data: \( D = \{D_1, D_2, ..., D_n\} \) a set of \( n \) examples
\( D_i = \langle \mathbf{x}_i, y_i \rangle \)
\( \mathbf{x}_i = (x_{i,1}, x_{i,2}, \cdots x_{i,d}) \) is an input vector of size \( d \)
\( y_i \) is the desired output (given by a teacher)

Objective: learn the mapping \( f : X \rightarrow Y \)
\( \) s.t. \( y_i \approx f(x_i) \) for all \( i = 1, \ldots, n \)

- **Regression**: \( Y \) is continuous
  Example: earnings, product orders \( \rightarrow \) company stock price
- **Classification**: \( Y \) is discrete
  Example: handwritten digit in binary form \( \rightarrow \) digit label

Linear regression

- **Function** \( f : X \rightarrow Y \) is a linear combination of input components
\[
f(x) = w_0 + w_1x_1 + w_2x_2 + \ldots w_dx_d = w_0 + \sum_{j=1}^{d} w_jx_j
\]
\( w_0, w_1, \ldots w_k \) - parameters (weights)

Bias term \( \rightarrow 1 \)

Input vector \( \mathbf{x} \)
\[
\sum f(x, \mathbf{w})
\]

```
- \( \)  

\[
\]
Linear regression

• **Shorter (vector) definition of the model**
  – Include bias constant in the input vector
    \[ x = (1, x_1, x_2, \ldots, x_d) \]
    \[ f(x) = w_0 x_0 + w_1 x_1 + w_2 x_2 + \ldots + w_d x_d = w^T x \]
    \[ w_0, w_1, \ldots, w_k \text{ - parameters (weights)} \]

\[ (1, x_1, x_2, \ldots, x_d) \]

\[ f(x, w) \]

\[ \sum \]

Input vector

\[ x \]

\[ x_1 \]

\[ x_2 \]

\[ \vdots \]

\[ x_d \]

\[ w_0 \]

\[ w_1 \]

\[ w_2 \]

\[ w_d \]

\[ f(x, w) \]

---

**Linear regression. Error.**

• **Data:** \( D_i = \langle x_i, y_i \rangle \)
• **Function:** \( x_i \rightarrow f(x_i) \)
• **We would like to have** \( y_i \approx f(x_i) \text{ for all } i = 1, \ldots, n \)

• **Error function**
  – measures how much our predictions deviate from the desired answers

\[ J_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 \]

• **Learning:**
  We want to find the weights minimizing the error!
Linear regression. Example

- 1 dimensional input \( x = (x_1) \)

- 2 dimensional input \( x = (x_1, x_2) \)
Linear regression. Optimization.

• We want the **weights minimizing the error**

\[ J_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i)^2 \]

• For the optimal set of parameters, derivatives of the error with respect to each parameter must be 0

\[ \frac{\partial}{\partial w_j} J_n(w) = -2 \frac{1}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) x_{i,j} = 0 \]

• Vector of derivatives:

\[ \text{grad}_w (J_n(w)) = \nabla_w (J_n(w)) = -2 \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i) x_i = \bar{0} \]

---

Linear regression. Optimization.

• \( \text{grad}_w (J_n(w)) = \bar{0} \) defines a set of equations in \( w \)

\[ \frac{\partial}{\partial w_j} J_n(w) = -2 \frac{1}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) x_{i,j} = 0 \]

\[ \frac{\partial}{\partial w_0} J_n(w) = \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) = 0 \]

\[ \frac{\partial}{\partial w_1} J_n(w) = \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) = 0 \]

\[ \ldots \]

\[ \frac{\partial}{\partial w_d} J_n(w) = \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) = 0 \]

---
Solving linear regression

\[ \nabla \frac{\partial}{\partial w_j} J_n(w) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w_0 x_{i,0} - w_1 x_{i,1} - \ldots - w_d x_{i,d}) x_{i,j} = 0 \]

By rearranging the terms we get a \textit{system of linear equations} with \(d+1\) unknowns

\[ \mathbf{A} \mathbf{w} = \mathbf{b} \]

\[
\begin{align*}
    w_0 \sum_{i=1}^{n} x_{i,0} l + w_l \sum_{i=1}^{n} x_{i,1} l + \ldots + w_j \sum_{i=1}^{n} x_{i,j} l + \ldots + w_d \sum_{i=1}^{n} x_{i,d} l &= \sum_{i=1}^{n} y_i l \\
    w_0 \sum_{i=1}^{n} x_{i,0} x_{i,1} + w_l \sum_{i=1}^{n} x_{i,1} x_{i,1} + \ldots + w_j \sum_{i=1}^{n} x_{i,j} x_{i,j} + \ldots + w_d \sum_{i=1}^{n} x_{i,d} x_{i,1} &= \sum_{i=1}^{n} y_i x_{i,1} \\
    \vdots & \\
    w_0 \sum_{i=1}^{n} x_{i,0} x_{i,j} + w_l \sum_{i=1}^{n} x_{i,1} x_{i,j} + \ldots + w_j \sum_{i=1}^{n} x_{i,j} x_{i,j} + \ldots + w_d \sum_{i=1}^{n} x_{i,d} x_{i,j} &= \sum_{i=1}^{n} y_i x_{i,j}
\end{align*}
\]

\[ \mathbf{A} \mathbf{w} = \mathbf{b} \]

Solving linear regression

- The optimal set of weights satisfies:
  \[
  \nabla_w (J_n(w)) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - \mathbf{w}^T \mathbf{x}_i) \mathbf{x}_i = \mathbf{0}
  \]
  Leads to a \textit{system of linear equations (SLE)} with \(d+1\) unknowns of the form
  \[ \mathbf{A} \mathbf{w} = \mathbf{b} \]

\[
\begin{align*}
    w_0 \sum_{i=1}^{n} x_{i,0} x_{i,j} + w_l \sum_{i=1}^{n} x_{i,1} x_{i,j} + \ldots + w_j \sum_{i=1}^{n} x_{i,j} x_{i,j} + \ldots + w_d \sum_{i=1}^{n} x_{i,d} x_{i,j} &= \sum_{i=1}^{n} y_i x_{i,j}
\end{align*}
\]

\textit{Solution to SLE: ?}
Solving linear regression

- The optimal set of weights satisfies:
  \[ \nabla_w (J_n (w)) = -\frac{2}{n} \sum_{i=1}^{n} (y_i - w^T x_i) x_i = \mathbf{0} \]
  Leads to a **system of linear equations (SLE)** with \( d+1 \) unknowns of the form
  \[ \mathbf{A} \mathbf{w} = \mathbf{b} \]

\[
\begin{align*}
    w_0 \sum_{i=1}^{n} x_{i,0} x_{i,j} + w_1 \sum_{i=1}^{n} x_{i,1} x_{i,j} + \ldots + w_j \sum_{i=1}^{n} x_{i,j} x_{i,j} + \ldots + w_d \sum_{i=1}^{n} x_{i,d} x_{i,j} = \sum_{i=1}^{n} y_i x_{i,j}
\end{align*}
\]

**Solution to SLE:**
\[ \mathbf{w} = \mathbf{A}^{-1} \mathbf{b} \]

- matrix inversion