EE269
Signal Processing for Machine Learning

Lecture 12

Instructor : Mert Pilanci

Stanford University

February 25, 2019

Least Squares Regression

- Predict the value of a continuous target variable y
  $(x_1, y_1), ..., (x_n, y_n)$
  $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$
- Linear regression $f(x) = w^T x + w_0 = \sum_{n=1}^{d} x[n]w[n] + w_0$

Least Squares Regression

- Predict the value of a continuous target variable y
  $(x_1, y_1), ..., (x_n, y_n)$
  $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$
- Linear regression $f(x) = w^T x + w_0 = \sum_{n=1}^{d} x[n]w[n] + w_0$
- Performance measure: minimum mean squared error

$$R(w, w_0) = \mathbb{E}_{x,y}\left[(f(x) - y)^2\right]$$

$P_{x,y}$ is not known, estimate risk directly

$$\min_{w,w_0} \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i - w_0)^2$$

Least Squares Regression

- Predict the value of a continuous target variable y
  $(x_1, y_1), ..., (x_n, y_n)$
  $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$
- Linear regression $f(x) = w^T x + w_0 = \sum_{n=1}^{d} x[n]w[n] + w_0$
- Performance measure: minimum mean squared error

$$R(w, w_0) = \mathbb{E}_{x,y}\left[(f(x) - y)^2\right]$$

$P_{x,y}$ is not known, estimate risk directly

$$\min_{w,w_0} \; \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - w_0)^2$$

- add a regularization term $\lambda||w||_2^2$

$$\min_{w,w_0} \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - w_0)^2 + \lambda||w||_2^2$$

Least Squares Regression

- Loss function:
  $L(w, w_0) = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i - w_0)^2 + \lambda ||w||_2^2$
- $\frac{\partial}{\partial w_0} L(w, w_0) =$
  optimal $w_0^* = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i) = \bar{y} - w^T \bar{x}$
  where $\bar{x} = \sum_{i=1}^{n} x_i$ and $\bar{y} = \sum_{i=1}^{n} y_i$
- plugging w_0^* in $L(w, w_0)$
  $L(w, w_0^*) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y} - w^T (x_i - \bar{x}))^2 + \lambda ||w||_2^2$

Least Squares Regression

- Loss function:
  $L(w, w_0) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - w_0)^2 + \lambda||w||_2^2$
- $\frac{\partial}{\partial w_0} L(w, w_0) =$
  optimal $w_0^* = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i) = \bar{y} - w^T \bar{x}$
  where $\bar{x} = \sum_{i=1}^{n} x_i$ and $\bar{y} = \sum_{i=1}^{n} y_i$
- plugging w_0^* in $L(w, w_0)$
  $L(w, w_0^*) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y} - w^T(x_i - \bar{x}))^2 + \lambda||w||_2^2$
  define centered signals $\tilde{x} = x - \bar{x}$ and $\tilde{y} = y - \bar{y}$

$$\min_{w} \; ||\tilde{X}w - \tilde{y}||_2^2 + n\lambda||w||_2^2$$

Least Squares Regression

- Loss function:
  $L(w, w_0) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - w_0)^2 + \lambda||w||_2^2$
- $\frac{\partial}{\partial w_0}L(w, w_0) =$
  optimal $w_0^* = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i) = \bar{y} - w^T\bar{x}$
  where $\bar{x} = \sum_{i=1}^{n} x_i$ and $\bar{y} = \sum_{i=1}^{n} y_i$
- plugging w_0^* in $L(w, w_0)$
  $L(w, w_0^*) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y} - w^T(x_i - \bar{x}))^2 + \lambda||w||_2^2$
  define centered signals $\tilde{x} = x - \bar{x}$ and $\tilde{y} = y - \bar{y}$

$$\min_{w} \; ||\tilde{X}w - \tilde{y}||_2^2 + n\lambda||w||_2^2$$

$\frac{\partial}{\partial w}L(w, w_0^*) = 2\tilde{X}^T(\tilde{X}w^* - \tilde{y}) + 2n\lambda w^* = 0$

optimal solution $w^* = (\tilde{X}^T\tilde{X} + n\lambda I)^{-1}\tilde{X}^T\tilde{y}$

Least Squares Regression: Prediction

- Loss function:
  $L(w, w_0) = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i - w_0)^2 + \lambda ||w||_2^2$
  define centered signals $\tilde{x} = x - \bar{x}$ and $\tilde{y} = y - \bar{y}$
  optimal $w_0^* = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i) = \bar{y} - {w^*}^T \bar{x}$
  optimal solution $w^* = (\tilde{X}^T \tilde{X} + n\lambda I)^{-1} \tilde{X}^T \tilde{y}$
  Given a test signal x, the prediction is $f(x) = w_0^* + {w^*}^T x$

Least Squares Regression: Prediction

- Loss function:
  $L(w, w_0) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - w_0)^2 + \lambda||w||_2^2$
  define centered signals $\tilde{x} = x - \bar{x}$ and $\tilde{y} = y - \bar{y}$
  optimal $w_0^* = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i) = \bar{y} - {w^*}^T \bar{x}$
  optimal solution $w^* = (\tilde{X}^T \tilde{X} + n\lambda I)^{-1} \tilde{X}^T \tilde{y}$
  Given a test signal x, the prediction is $f(x) = w_0^* + {w^*}^T x$

$$
\begin{aligned}
\hat{f}(x) &= w_0^* + {w^*}^T x \\
&= \bar{y} - {w^*}^T x + {w^*}^T x \\
&= \bar{y} + {w^*}^T (x - \bar{x})
\end{aligned}
$$

Autoregressive models

- Predict current sample from the past using a weighted average

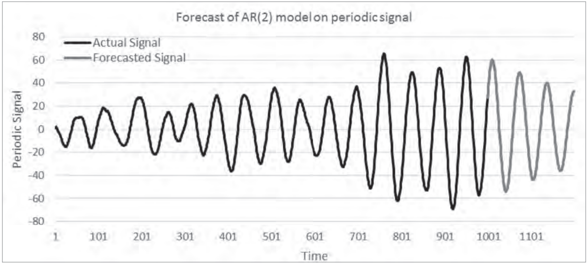
$$x[n] = \sum_k w_k x[n-k] + e_n$$

- e_t is an error term
- Matrix vector form $x = Aw + e$
- Least squares optimization problem $w^* = \arg\min ||Aw - x||_2^2$

Autoregressive models: forecasting

- Predict current sample from the past using a weighted average

$$x[n] = \sum_k w_k x[n-k] + e_n$$



Sinusoids

$$x[n] = \sum_k w_k x[n-k] + e_n$$

- AR(2) model : two non-zero filter coefficients

$$x[n+1] = -w_0 x[n] - w_1 x[n-1]$$

  and error term $e_n = 0$
- Example: Sine wave $x[n] = \sin(\alpha n)$

Sinusoids

$$x[n] = \sum_k w_k x[n-k] + e_n$$

- AR(2) model : two non-zero filter coefficients

  $$x[n+1] = -w_0 x[n] - w_1 x[n-1]$$

  and error term $e_n = 0$
- Example: Sine wave $x[n] = \sin(\alpha n)$

  Recall $\sin(a+b) + \sin(a-b) = 2\cos(b)\sin(a)$

Sinusoids

$$x[n] = \sum_k w_k x[n-k] + e_n$$

- AR(2) model : two non-zero filter coefficients

$$x[n+1] = -w_0 x[n] - w_1 x[n-1]$$

  and error term $e_n = 0$
- Example: Sine wave $x[n] = \sin(\alpha n)$

  Recall $\sin(a+b) + \sin(a-b) = 2\cos(b)\sin(a)$

$$\begin{aligned} x[n+1] &= \sin(\alpha(n+1)) = \sin(\alpha n + \alpha) \\ &= -\sin(\alpha(n-1)) - 2\cos(\alpha)\sin(\alpha n) \\ &= -x[n-1] - 2\cos(\alpha) x[n] \end{aligned}$$

Autoregressive models: predicting missing samples

- After fitting the autoregressive model, we have a linear system of equations in $x[n]$

$$x[n] = \sum_k w_k x[n-k]$$

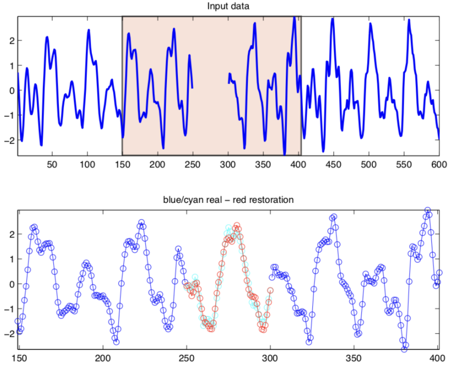
$$\begin{bmatrix} -w_p & \dots & -w_1 & 1 & 0 & 0 & & 0 \\ 0 & -w_p & \dots & -w_1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -w_p & \dots & -w_1 & 1 \end{bmatrix} x = 0$$

- we can solve for missing samples, e.g., $x[m_1], ...x[m_2]$

Autoregressive models: predicting missing samples

- After fitting the autoregressive model, we can predict unseen values

$$x[n] = \sum_k w_k x[n-k]$$



slide credit: P. Smaragdis

Autoregressive models: predicting missing samples

- After fitting the autoregressive model, we can predict unseen values

$$x[n] = \sum_k w_k x[n-k]$$