$w^{(n-1)}$.

$(n-1)\times p$.

$$\begin{bmatrix} \\ \\ \end{bmatrix} \quad \begin{bmatrix} \\ \\ \end{bmatrix} y_{n-1}$$

X_{n-1}.

$$w^{(n-1)} = (X_{n-1}^T X_{n-1})^{-1} X_{n-1}^T y_{n-1}. \Rightarrow p\times 1.$$

$(n\times p)$

$$X_n = \begin{bmatrix} X_{n-1} \\ \hline x_n \end{bmatrix} \in \mathbb{R}^{n\times p}.$$

$(n-1+1)$

$(p\times p)$

$R_{n-1} = X_{n-1}^t X_{n-1} \leftarrow n-1$ data

$R_n = X_n^t X_n \leftarrow n$ data

$$R_n = R_{n-1} + x_n x_n^t$$

$(p\times p)$ $\quad$ $p\times p.$ $\quad$ $(p\times 1 \cdot 1\times p)$

rank one update.

$$(A + xx^t)^{-1} = A^{-1} - \frac{A^{-1}xx^tA^{-1}}{1+x^tA^{-1}x}$$

$R_n^{-1} x_n^t y \quad \in \mathbb{R}^{p\times 1}$

$Z_n = x_n^t y_n. \in \mathbb{R}^{p}\times 1$

goal:

time n: $\boxed{w^{(n)} = R_n^{-1} Z_n.}$ from $n-1$ $\leftarrow w^{n-1}$

Error: $e_{n-1} = w^T x_{n-1} - y_{n-1}$

(each data)

$$\min \sum_{i=0}^{n-1} \lambda^{(n-1)-i} e_i^2. \ \text{(forgetting past)}$$

$$R_{n-1} = \sum_{i=0}^{n-1} \lambda^{(n-1)-i} x_i x_i^T \quad (p\times p)$$

$$Z_{n-1} = \sum_{i=0}^{n-1} \lambda^{(n-1)-i} x_i y_i \quad (p\times 1)$$

$$R_n = \sum_{i=0}^{n-1} \lambda^{(n-1)} x_i x_i^T + x_n x_n^T.$$

$$\left(\sum_{i=0}^{N} \lambda^{n-i} x_i x_i^T\right).$$

$$= \lambda\left[\sum_{i=0}^{n-1} \lambda^{(n-1)-i} x_i x_i^T\right] + x_n x_n^T$$

$$= \lambda R_{n-1} + x_n x_n^T$$

rank one update

$$\left(A^{-1} - \frac{A^{-1}xx^tA^{-1}}{1+x^tA^{-1}x}\right)(A + xx^t)$$

$$+ A^{-1}xx^tA^{-1}xx^t$$

$$= \frac{\cdots}{1+x^tA^{-1}x}$$

$$= I + \frac{A^{-1}xx^t\overbrace{x^tA^{-1}x}^{1\times 1} - A^{-1}x\overbrace{x^tA^{-1}x}^{1\times 1}x^t}{1+x^tA^{-1}x}$$

$$= I.$$

$$R_n^{-1} = \frac{1}{\lambda}R_{n-1}^{-1} - \frac{\frac{1}{\lambda^2}R_{n-1}^{-1}\, x_n x_n^t \, R_{n-1}^{-1}}{1+x_n^t\frac{1}{\lambda}R_{n-1}^{-1}x_n}$$

Define: $P_n = R_n^{-1}$, $P_{n-1} = R_{n-1}^{-1}$

$$P_n = \frac{1}{\lambda}P_{n-1} - \frac{\frac{1}{\lambda^2}P_{n-1}\, x_n x_n^t P_{n-1}}{1+x_n^T\frac{1}{\lambda}P_{n-1}x_n} = \frac{1}{\lambda}P_{n-1} - \frac{1}{\lambda}k_n x_n^T P_{n-1}$$

$$k_n = \frac{\frac{1}{\lambda}P_{n-1}x_n}{1+x_n^t\frac{1}{\lambda}P_{n-1}x_n}$$

$$z_n = \sum_{i=0}^{n}\lambda^{n-i}x_i y_i = \sum_{i=0}^{n-1}\lambda^{n-i}x_i y_i + x_n y_n.$$

$$= \lambda\sum_{i=0}^{n-1}\lambda^{n-1-i}x_i y_i + x_n y_n$$

$$= \lambda z_{n-1} + x_n y_n$$

homework

$$k_n = \left[\frac{1}{\lambda}P_{n-1} - \frac{1}{\lambda}k_n x_n^T P_{n-1}\right]x_n = P_n x_n$$

$$k_n + k_n x_n^t \frac{1}{\lambda}P_{n-1}x_n = \frac{1}{\lambda}P_{n-1}x_n.$$

$$k_n = \frac{1}{\lambda}P_{n-1}x_n - \frac{1}{\lambda}k_n x_n^t P_{n-1}x_n.$$

Goal: update for w at time n

$$w^n = P_n z_n = P_n[\lambda z_{n-1} + x_n y_n] = \lambda P_n z_{n-1} + P_n x_n y_n.$$

$$= \lambda\left[\frac{1}{\lambda}P_{n-1} - \frac{1}{\lambda}k_n x_n^T P_{n-1}\right] z_{n-1} + P_n x_n y_n$$

$$= \underbrace{P_{n-1} z_{n-1}}_{(w^{n-1})} - k_n x_n^T \overbrace{(P_{n-1} z_{n-1})}^{w^{n-1}} + \overbrace{(P_n x_n)}^{k_n} y_n.$$

$$= w^{n-1} - k_n(x_n^T w^{(n-1)} - y_n)$$

Gain × Prediction error.

Starting Point: Linear Regression

- Data: $\{\mathbf{x}_n, y_n\}_{n=1}^N$, $\mathbf{x} \in \mathcal{R}^p$, $y_n \in \mathcal{R}$
- Usually $N > p$, Offset / bias term absorbed into $\mathbf{w}$.
- Model: $y = \mathbf{w}^T \mathbf{x}$
- Linear model on fixed nonlinear basis functions:

$$y = \mathbf{w}^T \mathbf{\Phi}(\mathbf{x})$$

  where $\mathbf{\Phi}(.)$ is from a set of fixed transforms *e.g.* Radial Basis Functions (RBF), with some hyperparameters in them.
- We minimize error: $E = ||\mathbf{y} - X\mathbf{w}||^2$
- X is an $N \times p$ matrix; $\mathbf{y}$ is $N \times 1$ vector.
- There is a closed form solution: $\mathbf{w} = (X^t X)^{-1} X^t \mathbf{y}$
- We can estimate $\mathbf{w}$ by a gradient descent algorithm:

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \eta \nabla_{\mathbf{w}} E$$

- $\nabla_{\mathbf{w}} E$ is a vector, dimension p, derivative of E with respect to each of the weights:

$$\nabla_{\mathbf{w}} E = 2X^t(\mathbf{y} - X\mathbf{w})$$ (少负号).

- This gradient is sum over all the data; we could also perform sample by sample update: Stochastic Gradient Descent

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \eta(y(n) - \mathbf{w}^t \mathbf{x}_n)\mathbf{x}_n$$

- Structure: Constant × Error × Input

Introduction

In online estimation of a regression problem defined by $\{\boldsymbol{x}_n, y_n\}_{n=1}^N$ where the inputs $\boldsymbol{x}_n \in \mathcal{R}^p$ and the target y_n is scalar, we consider the arrival of data to be sequential

$$\ldots\ \{\boldsymbol{x}_{n-1}, y_{n-1}\},\ \{\boldsymbol{x}_n, y_n\},\ \{\boldsymbol{x}_{n+1}, y_{n+1}\},\ \ldots$$

Our task is to update the solution for the parameters $\boldsymbol{w}_{n-1}$ we have at time $(n-1)$ upon the arrival of data $\{\boldsymbol{x}_n, y_n\}$ at time n *without* refering back to data seen in the past: $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n-1}$. What summary information do we carry with us is the basic question underlying this topic. The RLS algorithm achieves this by updating an inverse of a matrix (inverse covariance matrix, P_{n-1}) sequentially. The algorithm is as follows:

设 P_{n-1} 初值.

$$\begin{aligned}
\boldsymbol{k}_n &= \frac{\frac{1}{\lambda} P_{n-1}\, \boldsymbol{x}_n}{1 + \frac{1}{\lambda} \boldsymbol{x}_n^T P_{n-1}\, \boldsymbol{x}_n} \\
e(n) &= y_n - \boldsymbol{w}_{n-1}^T \boldsymbol{x}_n \\
\boldsymbol{w}_n &= \boldsymbol{w}_{n-1} + \boldsymbol{k}_n\, e(n) \\
P_n &= \frac{1}{\lambda} P_{n-1} - \frac{1}{\lambda} \boldsymbol{k}_n\, \boldsymbol{x}_n^T P_{n-1}
\end{aligned}$$

Note several aspects of the algorithm:

- There is no matrix inversion at each step, the update of P_{n-1} uses the matrix inversion lemma.
- There is a gain term $\boldsymbol{k}_n$ which gives a structure: the new values of parameters being the old values to which we add a term given by the error and this gain.
- $0 < \lambda < 1$ is a user defined parameter that controls how fast the contribution of data seen in the past decays.

伪代码:

$P_{n-1} = I/C$ $(X[0])$

w_{n-1} = np.zeros((P,1))

for n in range(N):

$k_n = \frac{1}{\lambda} P_{n-1} x_n / 1 + \frac{1}{\lambda} x_n^T P_{n-1} x_n$.

$e(n) = y_n - w_{n-1}^T x_n$.

$w_n = w_{n-1} + k_n e(n)$

$P_n = \frac{1}{\lambda} P_{n-1} - \frac{1}{\lambda} k_n x_n^T P_{n-1}$

更新 P_{n-1}, w_{n-1}