Math 1: Linear Algebra

Eric@Udacity

Content

- Linear Algebra

  (1) Vector and Matrix

  (2) Solve Linear equations

  (3) Eigenvalues and Eigenvectors

  (4) Singular Value Decomposition (LSA)

1.1 Vector and Matrix

- Scalar, vector, matrix, tensor
- Operator: +, -, $\times$, T, ·, inverse (A-1), determinant (|A|), adjoint matrix (A*)

  (1) dot

  (2) inverse

  (3) determinant

  (4) adjoint matrix

- **Terms related to matrix**

  (1) Order of matrix ($m \times n$)

  (2) Square matrix

  (3) Diagonal matrix

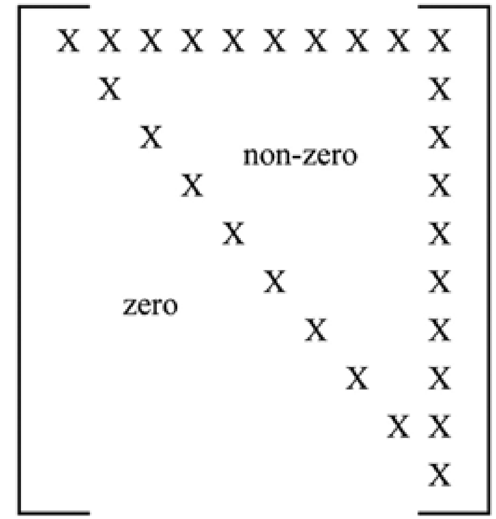
  (4) Upper triangular matrix

  (5) Lower triangular matrix

  (6) Scalar matrix: Square matrix with all the diagonal elements equal to some constant k.

  (7) Identity matrix

  A*I = A



(8) Trace: the sum of all the diagonal elements of a square matrix

(9) Symmetrical matrix A = A.T

(10) Orthogonal matrix A.inv = A.T, A dot A.T = I

(11) Rank of matrix

(12) Nonsingular matrix Note: A is |A| !=0 (A is a square matrix)

(13) p-norm

$$\|x\|_p = (\sum_i |x_i|^p)^{\frac{1}{p}} \quad \|x\|^\infty = \max_i |x_i|$$

(14) Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i,j} |A_{i,j}|^2}$$

(15) Condition number (nonsingular)

Stability $\kappa(A) = \|A\|_F \|A^{-1}\|_F$

(1) Dot

2 × 3 dot 3 × n = 2 × n

$$\begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 + 1x_2 \\ 1x_1 + 5x_2 \end{bmatrix}$$

(2) Inverse (elementary transformation)

Row switching: A row within the matrix can be switched with another row.

Row multiplication: Each element in a row can be multiplied by a non-zero constant.

Row addition: A row can be replaced by the sum of that row and a multiple of another row.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 5 & 6 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{-5R_1 + R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & -4 & -15 & -5 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{4R_2 + R_3} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & 1 & -5 & 4 & 1 \end{array}\right]$$

$$\xrightarrow{-2R_2 + R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & -5 & 1 & -2 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & 1 & -5 & 4 & 1 \end{array}\right]$$

$$\underset{\xrightarrow{-4R_3 + R_2}}{\xrightarrow{5R_3 + R_1}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -24 & 18 & 5 \\ 0 & 1 & 0 & 20 & -15 & -4 \\ 0 & 0 & 1 & -5 & 4 & 1 \end{array}\right]$$

$$A^{-1} = \begin{bmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{bmatrix}$$

(3) Determinant

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

|A| = ad – bc

sum(-1**(i+j) *a_ij *det(sub_matrix))

(4) adjoint matrix A*

$$\boxed{A^{-1} = \frac{\text{Adj}\,A}{\det A}}$$

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 4 & 0 & 6 \\ 0 & 1 & -1 \end{bmatrix}$$

$$\text{cof}(a_{11}) = +\begin{vmatrix} 0 & 6 \\ 1 & -1 \end{vmatrix} = -6 \qquad \text{cof}(a_{12}) = -\begin{vmatrix} 4 & 6 \\ 0 & -1 \end{vmatrix} = 4$$

$$\text{cof}(a_{21}) = -\begin{vmatrix} -1 & 2 \\ 1 & -1 \end{vmatrix} = 1 \qquad \text{cof}(a_{22}) = +\begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix} = -1$$

$$\text{cof}(a_{31}) = +\begin{vmatrix} -1 & 2 \\ 0 & 6 \end{vmatrix} = -6 \qquad \text{cof}(a_{32}) = -\begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix} = 2$$

$$\text{cof}(a_{13}) = +\begin{vmatrix} 4 & 0 \\ 0 & 1 \end{vmatrix} = 4 \qquad \text{Adj}\,A = [\text{cof}(a_{ij})]^T = \begin{bmatrix} -6 & 4 & 4 \\ 1 & -1 & -1 \\ -6 & 2 & 4 \end{bmatrix}^T = \begin{bmatrix} -6 & 1 & -6 \\ 4 & -1 & 2 \\ 4 & -1 & 4 \end{bmatrix}$$

$$\text{cof}(a_{23}) = -\begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = -1$$

$$\text{cof}(a_{33}) = +\begin{vmatrix} 1 & -1 \\ 4 & 0 \end{vmatrix} = 4$$

1.2 Solve Linear equations

- OLS

For any questions like:

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \epsilon_i$$

where $i \in (1, 2, \ldots, n)$

We set y = $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $X = \begin{pmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \vdots & & & \\ 1 & x_{2n} & \cdots & x_{kn} \end{pmatrix}$ $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$ $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$

Thus, the function can be written as:

$$y = X\beta + \epsilon$$

We define the error as

$$S = \sum_i^n \epsilon_i^2 = \epsilon_i^T \epsilon_i = (y - X\beta)^T (y - X\beta) = y^T y - y^T X\beta - \beta X^T y + \beta^T X^T X\beta$$

Then we calculate the first order derivatives of the error term, we get

$$\frac{\partial S}{\partial \beta} = -2X^T y + 2X^T X\beta$$

Let the first order derivatives equals to zero, we can get

$$\beta = (X^T X)^{-1} X^T y$$

And we also need to check the second order derivatives

$$\frac{\partial^2 S}{\partial \beta^2} = 2X^T X$$

which is a positive matrix.

```python
class LinearReg:
    def __init__(self, data):
        '''data: type: pandas dataframe'''
        self.data = data
        self.length = len(data)
    def ols(self, x, y):
        '''
        x: column name
        y: column name
        '''
        X = np.matrix(np.vstack([np.ones(self.length), self.data[x].values]).T)
        y = np.matrix(self.data[y].values).T
        beta = np.linalg.inv(X.T*X)*X.T*y
        return beta
    def sklearn_ols(self, x, y):
        X = np.matrix(self.data[x].values).T
        y = np.matrix(self.data[y].values).T
        # Create linear regression object
        OLS = linear_model.LinearRegression()
        # Train the model using the training sets
        OLS.fit(X, y)
        return np.vstack([OLS.intercept_, OLS.coef_])
    def visual(self, x, y, step = 0.01):
        para = self.ols(x, y)
        X = self.data[x]
        Y = self.data[y]
        min_x, max_x = min(X), max(X)
        # x is also matrix
        func = lambda x: x*para
        x_sim = np.arange(min_x, max_x, step)
        xm = np.vstack([np.ones(len(x_sim)), x_sim]).T
        y_sim = func(xm)
        self.data.plot.scatter(x,y)
        plt.plot(x_sim, y_sim)
        plt.title('The Relationship Between {0} and {1}'.format(x, y))
```

- # Maximum Likelihood

First we drive the likelihood function:

$$p(\epsilon_i) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{\epsilon_i^2}{2\sigma^2})$$

$$p(y_i|x_i;\beta) = \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(y_i - \beta^T x_i)^2}{2\sigma^2})$$

$$L(\theta) = \prod_{i=1}^{n} p(y_i|x_i;\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(y_i - \beta^T x_i)^2}{2\sigma^2})$$

$$l(\theta) = logL(\theta) = log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(y_i - \beta^T x_i)^2}{2\sigma^2})$$

$$= \sum_{i=1}^{n} log \frac{1}{\sqrt{2\pi\sigma}} exp(-\frac{(y_i - \beta^T x_i)^2}{2\sigma^2})$$

$$= mlog \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \beta^T x_i)^2$$

Since $mlog\frac{1}{\sqrt{2\pi\sigma}}$ is a constant, so we only need to consider the minimum of

$$\frac{1}{2} \sum_{i=1}^{m} (y_i - \beta^T x_i)^2$$

We can also write above formula as:

$$\frac{1}{2}(X\theta - y)^T (X\theta - y)$$

Therefore, the following thing is almost like we write in problem 19, calculate the first order derivatives and let it equals to zero

The final result is also

$$\beta = (X^T X)^{-1} X^T y$$

1.3 Eigenvalues and Eigenvectors

If

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

then the characteristic equation is

$$\left|\mathbf{A} - \lambda \cdot \mathbf{I}\right| = \left|\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right| = 0$$

$$\left|\begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix}\right| = \lambda^2 + 3\lambda + 2 = 0$$

and the two eigenvalues are

$$\lambda_1=-1,\ \lambda_2=-2$$

All that's left is to find the two eigenvectors. Let's find the eigenvector, $\mathbf{v}_1$, associated with the eigenvalue, λ_1=-1, first.

$$\mathbf{A} \cdot \mathbf{v}_1 = \lambda_1 \cdot \mathbf{v}_1$$
$$(\mathbf{A} - \lambda_1) \cdot \mathbf{v}_1 = 0$$
$$\begin{bmatrix} -\lambda_1 & 1 \\ -2 & -3-\lambda_1 \end{bmatrix} \cdot \mathbf{v}_1 = 0$$
$$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \mathbf{v}_1 = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \begin{bmatrix} v_{1,1} \\ v_{1,2} \end{bmatrix} = 0$$

so clearly from the top row of the equations we get

$$v_{1,1} + v_{1,2} = 0, \quad \text{so}$$
$$v_{1,1} = -v_{1,2}$$

Note that if we took the second row we would get

$$-2 \cdot v_{1,1} + -2 \cdot v_{1,2} = 0, \quad \text{so again}$$
$$v_{1,1} = -v_{1,2}$$

In either case we find that the first eigenvector is any 2 element column vector in which the two elements have equal magnitude and opposite sign.

$$\mathbf{v}_1 = k_1 \begin{bmatrix} +1 \\ -1 \end{bmatrix}$$

where k_1 is an arbitrary constant. Note that we didn't have to use +1 and -1, we could have used any two quantities of equal magnitude and opposite sign.

Going through the same procedure for the second eigenvalue:

$$\mathbf{A} \cdot \mathbf{v}_2 = \lambda_2 \cdot \mathbf{v}_2$$

$$\left(\mathbf{A} - \lambda_2\right) \cdot \mathbf{v}_2 = \begin{bmatrix} -\lambda_2 & 1 \\ -2 & -3 - \lambda_2 \end{bmatrix} \cdot \mathbf{v}_2 = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_{2,1} \\ v_{2,2} \end{bmatrix} = 0 \quad \text{so}$$

$$2 \cdot v_{2,1} + 1 \cdot v_{2,2} = 0 \quad \left(\text{or from bottom line: } -2 \cdot v_{2,1} - 1 \cdot v_{2,2} = 0\right)$$

$$2 \cdot v_{2,1} = -v_{2,2}$$

$$\mathbf{v}_2 = k_2 \begin{bmatrix} +1 \\ -2 \end{bmatrix}$$

Again, the choice of +1 and -2 for the eigenvector was arbitrary; only their ratio is important. This is demonstrated in the MatLab code below.

1.4 Singular Value Decomposition