Линейный регрессионный анализ

Нормальные уравнения

Регуляризация

Шаг 1. Линейный регрессионный анализ. Нормальные уравнения.

$$y_i = \beta_0 + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \ldots + \beta_k \cdot x_{i,k} + \epsilon_i \quad i = 1, \ldots n \tag{1}$$

Будет удобнее записывать в матричном виде, если добавим фиктивную переменную, тождественно равную 1

$$x_{0,1} = x_{0,2} = \ldots = x_{0,n} = 1$$

$$y_i = \beta_0 \cdot x_{i,0} + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \ldots + \beta_k \cdot x_{i,k} + \epsilon_i \quad i = 1, \ldots n$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{1,0} & x_{1,1} & x_{1,2} & \ldots & x_{1,k} \\ x_{2,0} & x_{2,1} & x_{2,2} & \ldots & x_{1,k} \\ x_{3,0} & x_{3,1} & x_{3,2} & \ldots & x_{1,k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{n,0} & x_{n,1} & x_{n,2} & \ldots & x_{1,k} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \ldots \\ \epsilon_n \end{pmatrix}$$

В матричном виде

$$\mathbf{Y} = \mathbf{X} \cdot \vec{\beta} + \vec{\epsilon} \tag{2}$$

Критерий качества - СКО (SSE, MSE)

$$Q = \sum_{i=1}^{n} (y_i - (\beta_0 \cdot x_{i,0} + \beta_1 \cdot x_{i,1} + \beta_2 \cdot x_{i,2} + \ldots + \beta_k \cdot x_{i,k}))^2 \tag{3}$$

$$Q = \sum_{i=1}^{n} \epsilon_i^2$$

В матричном виде

$$Q = \vec{\epsilon}^T \cdot \vec{\epsilon} = (\mathbf{Y} - \mathbf{X} \cdot \vec{\beta})^T \cdot (\mathbf{Y} - \mathbf{X} \cdot \vec{\beta}) \tag{4}$$

Чтобы найти вектор $\vec{\beta}$, при котором критерий качества минимален, составляем систему уравнений

$$\{ \frac{\partial Q}{\partial \beta_i} = 0 \quad i = 0, \ldots, k$$

Систему уравнений, полученную после дифференцирования, запишем в матричном виде

$$-2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\vec{\beta} = \vec{0}$$

После преобразований получаем систему нормальных уравнений

$$\mathbf{X}^T\mathbf{X}\vec{\beta} = \mathbf{X}^T\mathbf{Y} \tag{5}$$

Обозначим $\hat{\beta}$ решение этого уравнения

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{6}$$

Шаг 2. Регуляризация

Меняем критерий качества.

$$Q = \sum_{i=1}^{n} \epsilon_i^2 + \lambda \cdot \sum_{i=1}^{k} \beta_i^2 \tag{7}$$

Цели регуляризации

1. Предотвратить чрезмерную подгонку (overfitting)
2. Включить в критерий качества штраф за сложность модели
3. Обеспечить существование обратной матрицы $(\mathbf{X}^T\mathbf{X})^{-1}$
4. Не допустить большие значения коэффициентов модели
5. ...

После дифференцирования измененного критерия качества получим систему уравнений

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X} + \lambda D)^{-1}\mathbf{X}^T\mathbf{Y} \tag{8}$$

где

$$D = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Как получить оценки параметров β методом скорейшего спуска
https://www.r-bloggers.com/machine-learning-ex-5-1-regularized-linear-regression/

Свободные члены не включаются в штраф.
Хоть в линейных регрессионных моделях, хоть в нейронных сетях

Как найти λ ?

Подобрать, чтобы минимизировать ошибку на тестовой выборке...

Три подвыборки: обучающая, валидации, тестовая.

Уменьшение дисперсии за счет увеличения смещения

Смещение - интерпретация коэффициентов регрессионного уравнения затруднена.

Другие способы штрафа больших значений весов

$$Q = \sum_{i=1}^{n} \epsilon_i^2 + \lambda \cdot \sum_{i=1}^{k} i\beta_i^2 \tag{9}$$

$$Q = \sum_{i=1}^{n} \epsilon_i^2 + \lambda \cdot \sum_{i=1}^{k} |\beta_i| \tag{10}$$

$$Q = \sum_{i=1}^{n} \epsilon_i^2 + \lambda \cdot \sum_{i=1}^{k} \beta_i^m \tag{11}$$

m < 1

Abu-Mostafa, Magdon-Ismail, Lin
Learning From Data
2012

Регуляризация: лекарство или отрава?

Регуляризация - необходимое зло. Ключево слово - необходимое.

Если наша модель слишком сложна для изучаемого набора данных, мы обречены. Но регуляризация дает нам шанс.