1. Методи попередження перенавчання для гаусового дискримінантного аналізу.

Існують два споріднених підхода до формалізації завдання навчання: перший заснований на введення функції втрат, другий - на введенні ймовірнісної моделі продовження роду даних. Обидва в підсумку призводять до схожих (іноді навіть в точності однаковим) оптимізаційних завдань.
Навчання - це оптимізація.

Перенавчання (у значенні "занадто", англ. overfitting) в машинному навчанні і статистиці - явище, коли побудована модель добре пояснює приклади з навчальної вибірки, але відносно погано працює на прикладах, які не брали участі у навчанні (на прикладах з тестової вибірки).

Це пов’язано з тим, що при побудові моделі ("у процесі навчання") в навчальній вибірці виявляються деякі випадкові закономірності, які відсутні в генеральній сукупності

Дискримінантний аналіз є близьким до дисперсійного і регресійного аналізів, які також намагаються виразити одну із залежних змінних у вигляді лінійної комбінації інших показників або вимірювань.Однак, у двох інших методів залежна змінна є числовий величиною, в той час як у дискримінантному аналізі це категорійна змінна.

Метод попередження перенавчання

- **Регуляризация параметров по норме L2**
  Цель такой стратегии регуляризации – выбирать веса, близкие к началу координат1, за счет прибавления к целевой функции члена регуляризации $\Omega(\theta) = 1/2 || w || 2$. В других науч-ных сообществах регуляризацию по норме L2 называют также гребневой регрессией, или регуляризацией Тихонова.
- **L1-регуляризация**
  Формально L1-регуляризация параметров модели w определяется по формуле

$$\Omega(\boldsymbol{\theta}) = \|\boldsymbol{w}\|_1 = \sum_i |w_i|,$$

т. е. как сумма абсолютных величин отдельных параметров1. Обсудим влияние L1-ре-гуляризации на простую модель линейной регрессии без параметра смещения – ту самую, что рассматривалась в ходе анализа L2-регуляризации. Особенно нас инте-ресуют различия между двумя видами регуляризации. Как и в предыдущем случае, сила регуляризации контролируется путем умножения штрафа на положительный гиперпараметр α. Таким образом, регуляризированная целевая функция J(w; X, y) описывается формулой

$$\tilde{J}(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}) = \alpha \|\boldsymbol{w}\|_1 + J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}),$$

где sign(w) означает, что функция sign применяется к каждому элементу w.

$$\nabla_{\boldsymbol{w}} \tilde{J}(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}) = \alpha \operatorname{sign}(\boldsymbol{w}) + \nabla_{\boldsymbol{w}} J(\boldsymbol{w}; \boldsymbol{X}, \boldsymbol{y}),$$

- **Штраф по норме как оптимизация с ограничениями**
- **Регуляризация и недоопределенные задачи**
- **Пополнение набора данных**
- **Робастность относительно шума**
- **Привнесение шума в выходные метки**
- **Обучение с частичным привлечением учителя**
- **Многозадачное обучение**
- **Ранняя остановка**

Обозначим $\boldsymbol{X}^{(\text{train})}$ и $\boldsymbol{y}^{(\text{train})}$ обучающий набор.
Разделить $\boldsymbol{X}^{(\text{train})}$ и $\boldsymbol{y}^{(\text{train})}$ на $(\boldsymbol{X}^{(\text{subtrain})}, \boldsymbol{X}^{(\text{valid})})$ и $(\boldsymbol{y}^{(\text{subtrain})}, \boldsymbol{y}^{(\text{valid})})$ соответственно.
Выполнить обучение с ранней остановкой (алгоритм 7.1), начав со случайных параметров $\boldsymbol{\theta}$ и используя $\boldsymbol{X}^{(\text{subtrain})}, \boldsymbol{y}^{(\text{subtrain})}$ в качестве обучающих данных, а $\boldsymbol{X}^{(\text{valid})}$ и $\boldsymbol{y}^{(\text{valid})}$ в качестве контрольных. В результате возвращается оптимальное число шагов i^*.
Снова присвоить $\boldsymbol{\theta}$ случайные значения.
Выполнить i^* шагов обучения на наборе $\boldsymbol{X}^{(\text{train})}, \boldsymbol{y}^{(\text{train})}$.

- **Связывание и разделение параметров**
- **итд**

2. В и р і ш и т и   с и с т е м у   л і н і й н и х   а р и ф м е т и ч н и х   р і в н я н ь:

(1) 8x1 + 2x2 - 3x3 + 42x4 = 1
(2) 13x1 + 13x2 = 4
(3) -7x1 + 11x3 = 10
(4) x2 + x4 = 23
(5) 2x1+x2-7x3 = (-4)

$$X := \begin{pmatrix} 8 & 2 & -3 & 42 \\ 13 & 13 & 0 & 0 \\ -7 & 0 & 11 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 1 & -7 & 0 \end{pmatrix} \qquad y := \begin{pmatrix} 1 \\ 4 \\ 10 \\ 23 \\ -4 \end{pmatrix}$$

w=(X^T * X)^-1 * X^T * y

$$X = \begin{pmatrix} 8 & 2 & -3 & 42 \\ 13 & 13 & 0 & 0 \\ -7 & 0 & 11 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 1 & -7 & 0 \end{pmatrix} \qquad X^T = \begin{pmatrix} 8 & 13 & -7 & 0 & 2 \\ 2 & 13 & 0 & 1 & 1 \\ -3 & 0 & 11 & 0 & -7 \\ 42 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$w := \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot y \qquad w = \begin{pmatrix} -3.07 \\ 3.45 \\ -0.682 \\ 0.407 \end{pmatrix}$$

$$X \cdot w = \begin{pmatrix} 1.456 \\ 4.935 \\ 13.994 \\ 3.857 \\ 2.081 \end{pmatrix} \qquad X \cdot w - y = \begin{pmatrix} 0.456 \\ 0.935 \\ 3.994 \\ -19.143 \\ 6.081 \end{pmatrix}$$