1. Метод наименьших квадратов

Пример 1. По обучающей выборке

x	-1	0	0	1	2
y	1	-2	1	7	8

методом наименьших квадратов построить модель вида $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.

Решение: Составляем матрицу $\mathbf{X}$ и вектор $\mathbf{y}$:

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 7 \\ 8 \end{pmatrix}.$$

Имеем

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 5 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{pmatrix}, \qquad \mathbf{X}^\top\mathbf{y} = \begin{pmatrix} 15 \\ 22 \\ 40 \end{pmatrix}.$$

Решая систему нормальных уравнений $\mathbf{X}^\top\mathbf{X}\beta = \mathbf{X}^\top\mathbf{y}$, находим

$$\beta = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

Таким образом, нашли модель $1 + 2x + x^2$ (синяя линия на графике).

Пример 2. Построить модель того же вида методом ридж-регрессии с параметром регуляризации $\lambda = 2$.

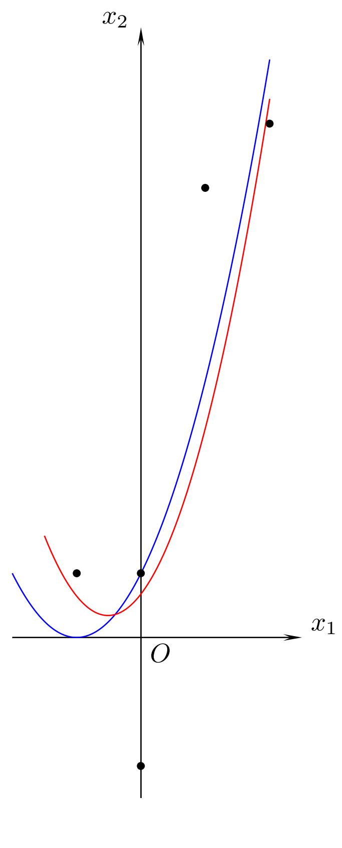
Решение: Для $\lambda = 2$ получаем

$$\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I} = \begin{pmatrix} 7 & 2 & 6 \\ 2 & 8 & 8 \\ 6 & 8 & 20 \end{pmatrix}.$$

Решая регуляризованную систему нормальных уравнений $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})\beta = \mathbf{X}^\top\mathbf{y}$, находим

$$\beta = \begin{pmatrix} {}^{21}\!/_{31} \\ {}^{81}\!/_{62} \\ {}^{79}\!/_{62} \end{pmatrix}.$$

Таким образом, нашли модель $\dfrac{21}{31} + \dfrac{81}{62}x + \dfrac{79}{62}x^2 = 0.6774 + 1.3065x + 1.2742x^2$ (красная линия на графике).



2. Дискриминантный анализ

Пример 3. Дана обучающая выборка:

x_1	0	2	1	1	1	2	4	4	4	6
x_2	4	4	3	5	4	1	0	2	1	1
y	0	0	0	0	0	1	1	1	1	1

1) Изобразить объекты обучающей выборки в пространстве признаков x_1, x_2;
2) с помощью линейного дискриминантного анализа для каждого класса построить дискриминантные функции и записать уравнение разделяющей поверхности; изобразить поверхность;
3) с помощью квадратичного дискриминантного анализа для каждого класса построить дискриминантные функции.

Решение: Оцениваем вероятности классов:

$$\widehat{\Pr}\{Y=0\} = \frac{1}{2}, \qquad \widehat{\Pr}\{Y=1\} = \frac{1}{2}.$$

Оцениваем средние для классов:

$$\widehat{\mu}_0 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \qquad \widehat{\mu}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

Выборочные матрицы ковариации для каждого класса:

$$\widehat{\mathbf{\Sigma}}_0 = \frac{1}{N_0-1}\sum_{y^{(i)}=0}(x^{(i)}-\widehat{\mu}_0)(x^{(i)}-\widehat{\mu}_0)^\top = \frac{1}{4}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \widehat{\mathbf{\Sigma}}_1 = \frac{1}{N_1-1}\sum_{y^{(i)}=1}(x^{(i)}-\widehat{\mu}_1)(x^{(i)}-\widehat{\mu}_1)^\top = \frac{1}{4}\begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}.$$

Оцениваем матрицу ковариации:

$$\widehat{\mathbf{\Sigma}} = \frac{1}{N-K}\sum_k\sum_{y^{(i)}=k}(x^{(i)}-\widehat{\mu}_k)(x^{(i)}-\widehat{\mu}_k)^\top = \frac{1}{8}\left(\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}\right) = \frac{1}{8}\begin{pmatrix} 10 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 5/4 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Находим

$$\widehat{\mathbf{\Sigma}}_0^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \widehat{\mathbf{\Sigma}}_1^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \widehat{\mathbf{\Sigma}}^{-1} = \begin{pmatrix} 4/5 & 0 \\ 0 & 2 \end{pmatrix}.$$

Линейные дискриминантные функции:

$$\delta_0(x) = x^\top\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mu}_0 - \frac{1}{2}\widehat{\mu}_0^\top\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mu}_0 + \ln\widehat{\Pr}\{Y=0\} = \frac{4}{5}x_1 + 8x_2 - \frac{82}{5} - \ln 2,$$

$$\delta_1(x) = x^\top\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mu}_1 - \frac{1}{2}\widehat{\mu}_1^\top\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mu}_1 + \ln\widehat{\Pr}\{Y=1\} = \frac{16}{5}x_1 + 2x_2 - \frac{37}{5} - \ln 2.$$

Разделяющая поверхность — прямая с уравнением $\delta_0(x) = \delta_1(x)$ (красная прямая на графике):

$$4x_1 - 10x_2 + 15 = 0.$$

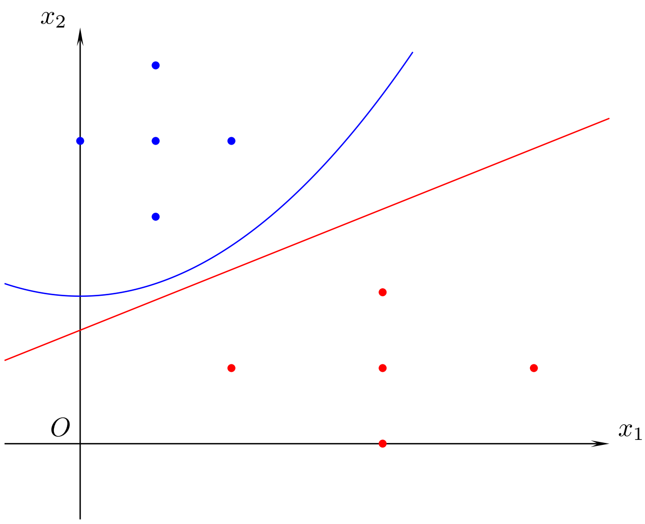
Квадратичные дискриминантные функции:

$$\delta_0(x) = -\frac{1}{2}\ln\det\widehat{\mathbf{\Sigma}}_0 - \frac{1}{2}(x-\widehat{\mu}_0)^\top\widehat{\mathbf{\Sigma}}_0^{-1}(x-\widehat{\mu}_0) + \ln\widehat{\Pr}\{Y=0\} = -x_1^2 + 2x_1 - x_2^2 + 8x_2 - 17,$$

$$\delta_1(x) = -\frac{1}{2}\ln\det\widehat{\mathbf{\Sigma}}_1 - \frac{1}{2}(x-\widehat{\mu}_1)^\top\widehat{\mathbf{\Sigma}}_1^{-1}(x-\widehat{\mu}_1) + \ln\widehat{\Pr}\{Y=1\} = -\frac{1}{4}x_1^2 + 2x_1 - x_2^2 + 2x_2 - 5 - \ln 2.$$

Разделяющая поверхность — парабола с уравнением (синяя кривая на графике)

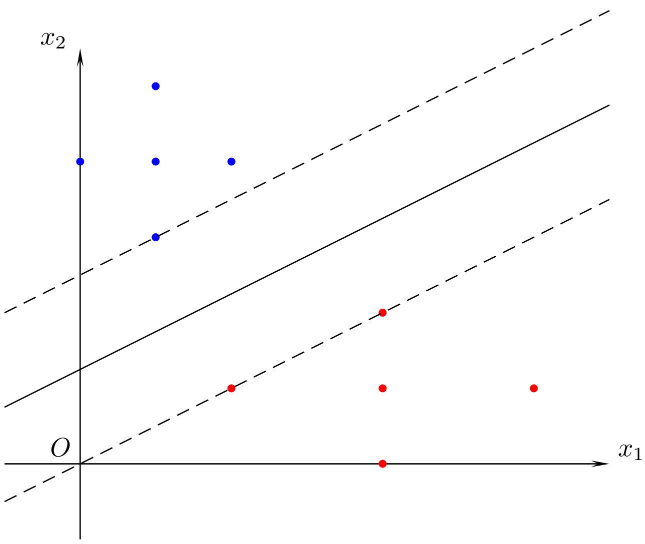
$$\frac{3}{4}x_1^2 - 6x_2 + 12 - \ln 2 = 0.$$



3. Оптимальная разделяющая гиперплоскость

Пример 4. Для данных из предыдущего примера найти уравнение оптимальной разделяющей гиперплоскости, указать опорные точки.

Решение: В качестве двух возможных наборов кандидатур в опорные точки подходят $(1,3)$, $(2,1)$, $(4,2)$ или $(1,3)$, $(2,4)$, $(2,1)$. Среди этих двух наборов нужно выбрать тот, для которого ширина «разделяющего коридора» больше. Для первого набора этот коридор заключен между прямыми (они изображены на рисунке — штриховая малиновая линия) $x_1 - 2x_2 = 0$ и $x_1 - 2x_2 = -5$. Расстояние между прямыми, т. е. ширина коридора, есть $\sqrt{5}$. Для второго набора точек получаем коридор, ограниченный прямыми $x_1 - x_2 = 1$, $x_1 - x_2 = -2$. Его ширина равна ${}^3/_{\sqrt{2}}$. Первый коридор шире, поэтому прямая $2x_1 - 4x_2 = -5$, проходящая через его центр, — оптимальная разделяющая, а точки $(1,3)$, $(2,1)$, $(4,2)$ — опорные.



4. Наивный байесовский классификатор

Пример 5. Дана обучающая выборка:

x_1	1	0	0	0	1	0	1	1	1	0
x_2	1	1	1	1	0	0	0	0	0	0
y	0	0	0	0	0	1	1	1	1	1

С помощью наивного байесова классификатора оценить апостериорные вероятности $\mathsf{Pr}(y\,|\,x)$, если (a) $x_1 = 1$, $x_2 = 0$; (b) $x_1 = 0$, $x_2 = 1$.

Оцениваем априорные вероятности:

$$\widehat{\mathsf{Pr}}\,\{Y=0\} = \frac{1}{2}, \qquad \widehat{\mathsf{Pr}}\,\{Y=1\} = \frac{1}{2}.$$

Оцениваем условные вероятности:

$$\widehat{\mathsf{Pr}}\,\{X_1=0\,|\,Y=0\} = \frac{3}{5}, \quad \widehat{\mathsf{Pr}}\,\{X_1=1\,|\,Y=0\} = \frac{2}{5}, \qquad \widehat{\mathsf{Pr}}\,\{X_1=0\,|\,Y=1\} = \frac{2}{5}, \quad \widehat{\mathsf{Pr}}\,\{X_1=1\,|\,Y=1\} = \frac{3}{5},$$

$$\widehat{\mathsf{Pr}}\left\{X_2 = 0 | Y = 0\right\} = \frac{1}{5}, \quad \widehat{\mathsf{Pr}}\left\{X_2 = 1 | Y = 0\right\} = \frac{4}{5}, \qquad \widehat{\mathsf{Pr}}\left\{X_2 = 0 | Y = 1\right\} = 1, \quad \widehat{\mathsf{Pr}}\left\{X_2 = 1 | Y = 1\right\} = 0.$$

Используя основное предположение наивного байесова классификатора, получаем

$$\mathsf{Pr}\left\{Y = 0 | X_1 = 1, X_2 = 0\right\} = \frac{\mathsf{Pr}\left\{X_1 = 1 | Y = 0\right\} \mathsf{Pr}\left\{X_2 = 0 | Y = 0\right\} \mathsf{Pr}\left\{Y = 0\right\}}{\mathsf{Pr}\left\{X_1 = 1,\, X_2 = 0\right\}} \approx \frac{\frac{2}{5}\cdot\frac{1}{5}\cdot\frac{1}{2}}{\frac{1}{25}+\frac{3}{10}} = \frac{\frac{1}{25}}{\frac{17}{50}} = \frac{2}{17},$$

$$\mathsf{Pr}\left\{Y = 1 | X_1 = 1, X_2 = 0\right\} = \frac{\mathsf{Pr}\left\{X_1 = 1 | Y = 1\right\} \mathsf{Pr}\left\{X_2 = 0 | Y = 1\right\} \mathsf{Pr}\left\{Y = 1\right\}}{\mathsf{Pr}\left\{X_1 = 1,\, X_2 = 0\right\}} \approx \frac{\frac{3}{5}\cdot 1\cdot\frac{1}{2}}{\frac{1}{25}+\frac{3}{10}} = \frac{\frac{3}{10}}{\frac{17}{50}} = \frac{15}{17}$$

(знаменатели в двух формулах выше равны сумме всех числителей и, следовательно окончательные оценки апостериорных вероятностей получаются после вычисления всех числителей).

$$\mathsf{Pr}\left\{Y = 0 | X_1 = 0, X_2 = 1\right\} = \frac{\mathsf{Pr}\left\{X_1 = 0 | Y = 0\right\} \mathsf{Pr}\left\{X_2 = 1 | Y = 0\right\} \mathsf{Pr}\left\{Y = 0\right\}}{\mathsf{Pr}\left\{X_1 = 0,\, X_2 = 1\right\}} \approx \frac{\frac{3}{5}\cdot\frac{4}{5}\cdot\frac{1}{2}}{\frac{6}{25}+0} = \frac{\frac{6}{25}}{\frac{6}{25}} = 1,$$

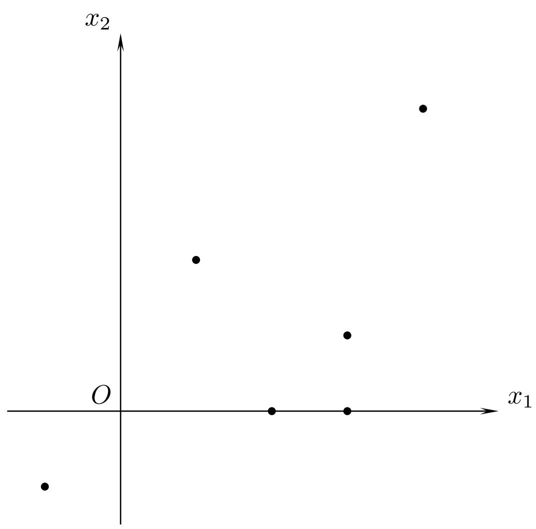
$$\mathsf{Pr}\left\{Y = 1 | X_1 = 0, X_2 = 1\right\} = \frac{\mathsf{Pr}\left\{X_1 = 0 | Y = 1\right\} \mathsf{Pr}\left\{X_2 = 1 | Y = 1\right\} \mathsf{Pr}\left\{Y = 1\right\}}{\mathsf{Pr}\left\{X_1 = 0,\, X_2 = 1\right\}} \approx \frac{\frac{2}{5}\cdot 0\cdot\frac{1}{2}}{\frac{6}{25}+0} = \frac{0}{\frac{6}{25}} = 0.$$

5. Метод главных компонент

Пример 6. Дана выборка:

x_1	4	-1	3	3	2	1
x_2	4	-1	0	1	0	2

Найти главные направления и объясненные дисперсии по главным компонентам.



Решение: Имеем

$$\mathbf{X} = \begin{pmatrix} 4 & 4 \\ -1 & -1 \\ 3 & 0 \\ 3 & 1 \\ 2 & 0 \\ 1 & 2 \end{pmatrix},$$

1 шаг. Центрируем данные. Находим $\overline{x} = (2,\, 1)$. Вычитая компоненты $\overline{x}$ из первого и второго столбца матрицы $\mathbf{X}$ соответственно, получаем

$$\mathbf{X}_c = \begin{pmatrix} 2 & 3 \\ -3 & -2 \\ 1 & -1 \\ 1 & 0 \\ 0 & -1 \\ -1 & 1 \end{pmatrix}.$$

2 шаг. Находим

$$\mathbf{C} = \mathbf{X}_c^\top \mathbf{X}_c = \begin{pmatrix} 16 & 10 \\ 10 & 16 \end{pmatrix}.$$

Матрица

$$\frac{1}{N-1}\mathbf{C} = \begin{pmatrix} 16/5 & 2 \\ 2 & 16/5 \end{pmatrix}.$$

— это выборочная матрица ковариации.

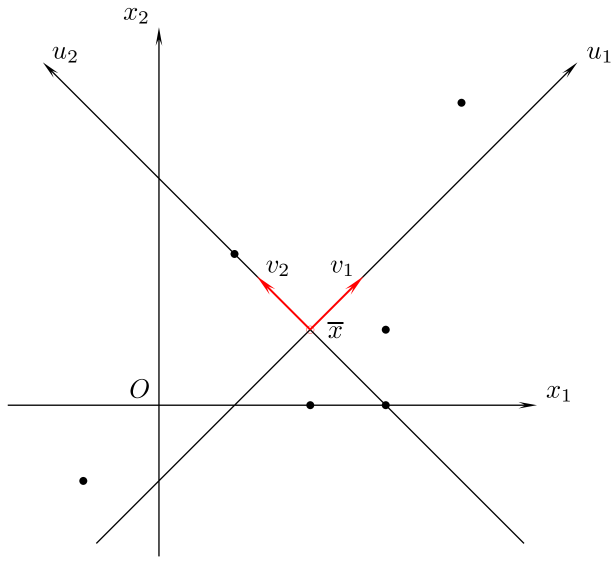
3 шаг. Находим собственные числа и собственные векторы матрицы $\mathbf{C}$:

$$\det(\mathbf{C} - \lambda \mathbf{I}) = \begin{vmatrix} 16-\lambda & 10 \\ 10 & 16-\lambda \end{vmatrix} = \lambda^2 - 32\lambda + 156 = (\lambda - 26)(\lambda - 6).$$

Собственные векторы нормируем:

$$\lambda_1 = 26, \quad v_1 = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \lambda_2 = 6, \quad v_2 = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Векторы v_1, v_2 — главные компоненты; значения $\lambda_1 = \sigma_1^2 = 26$, $\lambda_2 = \sigma_2^2 = 6$ — дисперсии по главным компонентам.



Найдем сингулярное разложение матрицы $\mathbf{X}_c = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \begin{pmatrix} \sqrt{26} & 0 \\ 0 & \sqrt{6} \end{pmatrix}, \qquad \mathbf{V} = (v_1 \mid v_2) = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{pmatrix};$$

$$\mathbf{U} = \mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 5\sqrt{13}/26 & \sqrt{3}/6 \\ -5\sqrt{13}/26 & \sqrt{3}/6 \\ 0 & -\sqrt{3}/3 \\ \sqrt{13}/26 & -\sqrt{3}/6 \\ -\sqrt{13}/26 & -\sqrt{3}/6 \\ 0 & \sqrt{3}/3 \end{pmatrix} = \begin{pmatrix} 0.6934 & 0.2887 \\ -0.6934 & 0.2887 \\ 0 & -0.5774 \\ 0.1387 & -0.2887 \\ -0.1387 & -0.2887 \\ 0 & 0.5774 \end{pmatrix}.$$

Напомним, что

- в матрице $\mathbf{V}$ по столбцам находятся векторы главных компонент v_1, v_2;
- в матрице $\boldsymbol{\Sigma}$ по диагонали — средние квадратические отклонения по главным компонентам σ_1, σ_2;
- в матрице $\mathbf{U}$ по строкам — координаты проекций точек на главные компоненты.