Линейная дискриминантная функция Фишера

- Данный подход не требует предположений о нормальном распределении данных.

- Пусть имеется обучающая выборка: x_1, x_2, ... x_N , где:

  N_1 элементов из множества X_1
  N_2 элементов из множества X_2

- Общее число элементов $N = N_1 + N_2$

- Задача состоит в построении разделяющей функции для двух классов:

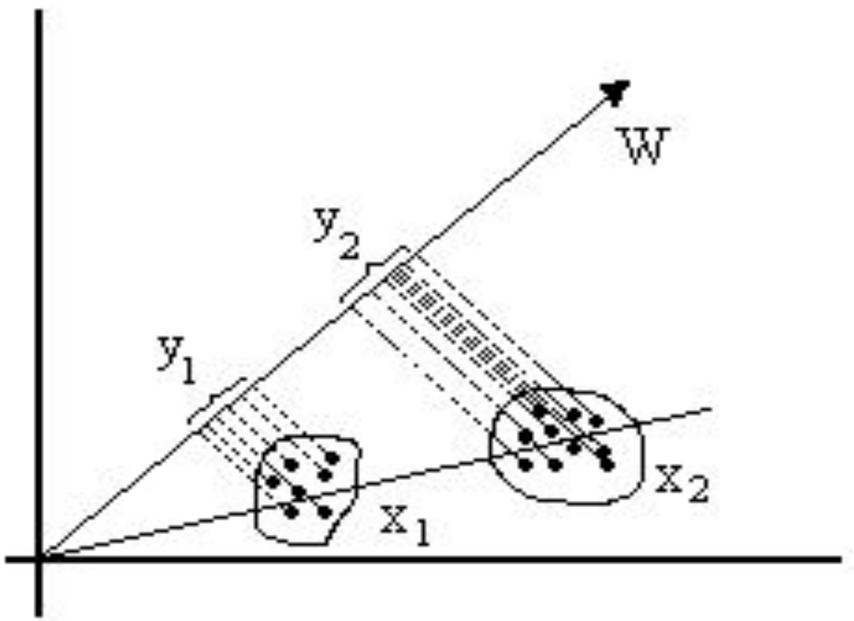
  $X_1 = \{x_i\}_{i=1..N1}$ $X_2 = \{\bar{x}_j\}_{j=1..N2}$

- Задача Фишера состоит в построении вырожденного линейного преобразования:

  $y = W^T \bar{x}$, $\|W\| = 1$

  $y = \|W\| \, \|x\| \cos(\widehat{\bar{W}, \bar{X}})$

  фактически это выражение дает нам проекции векторов X на вектор W
- У Фишера такое преобразование рассматривается как проекция на ось W: $\{\bar{x}\} \rightarrow \{y\} = Y$



- Нужно найти такой вектор W, чтобы множества Y_1 и Y_2 были наиболее разнесены (то есть, удалены друг от друга).

- Критерий разнесения может быть выбран разным.

- Исходить будем из следующих параметров: для каждой выборки определим среднее значение:

$$m_i = \frac{1}{N_i}\sum_{\chi\in\chi_i} x_i$$

$$\tilde{m}_i = \frac{1}{N_i}\sum_{y\in Y_i} y_i = \frac{1}{N_i}\sum_{x\in\chi_i} W^T x_i = W^T\{\frac{1}{N_i}\sum_{\chi\in\chi_i} x\} = W^T m_i$$

- Далее мы строим $|\tilde{m}_1 - \tilde{m}_2|$:

$|\tilde{m}_1 - \tilde{m}_2| = W^T(m_1 - m_2)$ - это скалярная величина

- Далее проблема состоит в оценке функции разброса:

$$\tilde{S}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2$$ - разброс внутри класса

$$\tilde{S}^2 = \tilde{S}_1^2 + \tilde{S}_2^2$$ - суммарный разброс

- Разброс внутри класса – это нечто вроде дисперсии, только ненормированной.

$$\frac{1}{N}\tilde{S}^2$$ - средняя дисперсия выборки в Y.

- Далее идет дело техники: как это вычислить и как оптимизировать.
- Мы определяем матрицу разброса внутри класса:

$$\tilde{S}_i^2 = \sum_{x\in\chi_i} (W^T x - W^T m_i)^2 = \sum_{x\in\chi_i} W^T (x - m_i)\, W^T (x - m_i) =$$

$$= \sum_{x\in\chi_i} W^T (x - m_i)(x - m_i)^T W = W^T \Big[\sum_{x\in\chi_i} (x - m_i)(x - m_i)^T\Big] W =$$

$$= W^T S_i W$$

S_i – матрица разброса внутри X_i.

- Таким образом получаем следующий результат:

$$\tilde{S}_i^2 = W^T S_i W$$

$S_1 + S_2 = S_W$ - суммарная матрица разброса для всех результатов.

$$\tilde{S}_1^2 + \tilde{S}_2^2 = W^T S_W W$$

- Таким же образом можно представить $(\tilde{m}_1 - \tilde{m}_2)^2$:

$$(\tilde{m}_1 - \tilde{m}_2)^2 = (W^T \bar{m}_1 - W^T \bar{m}_2)^2 = W^T(\bar{m}_1 - \bar{m}_2) W^T(\bar{m}_1 - \bar{m}_2) =$$

$$= W^T(\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T W = W^T S_B W$$

- Матрица S_B - матрица разброса между классами

- Тогда мы получаем искомый критерий в следующем виде:

$$J(\bar{W}) = \frac{W^T S_B W}{W^T S_W W}$$

- Далее стоит задача оптимизации данного отношения (мы должны его максимизировать).

- Рассмотрим свойства матрицы S_B:

$$S_B = (\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T$$

1. это квадратная матрица размерности n × n
2. произведение этой матрицы на произвольный вектор

$$S_B \bar{a} = (\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T \bar{a} = C(\bar{m}_1 - \bar{m}_2)$$

C - скаляр

дает вектор, который по направлению совпадает с разностью $(\bar{m}_1 - \bar{m}_2)$

3. ранг матрицы S_B равен единице, это легко показать, если представить ее в виде dd^T:

$$\left|\begin{matrix} d_1d_1 & d_1d_2 & \dots & d_1d_n \\ d_2d_1 & d_2d_2 & \dots & d_2d_n \\ . & & & \\ . & & & \\ d_nd_1 & d_nd_2 & \dots & d_nd_n \end{matrix}\right| = (\qquad) \cdot \bar{d}d_1 \; \bar{d}d_2 \; \bar{d}d_n$$

- матрица вырожденная и ранг ее равен единице.

Условная оптимизация по Лагранжу

- Запишем функцию Лагранжа: $F = W^T S_B W - \lambda W^T S_W W$,

  где λ -произвольная константа

  Эта функция зависит от W

  Мы должны найти производную этой функции по вектору :

$$\frac{dF}{d\overline{W}} = 2 S_B \overline{W} - 2 \lambda S_W \overline{W} = 0$$

$$\frac{d(X^T A X)}{d\overline{X}} = 2A\overline{X}$$ - такое правило существует, его легко доказать

- Получаем следующее:

  $S_B W - \lambda S_W W$

- C($\overline{m}_1$ - $\overline{m}_2$) = $\lambda S_W \overline{W}$, отсюда окончательно имеем:

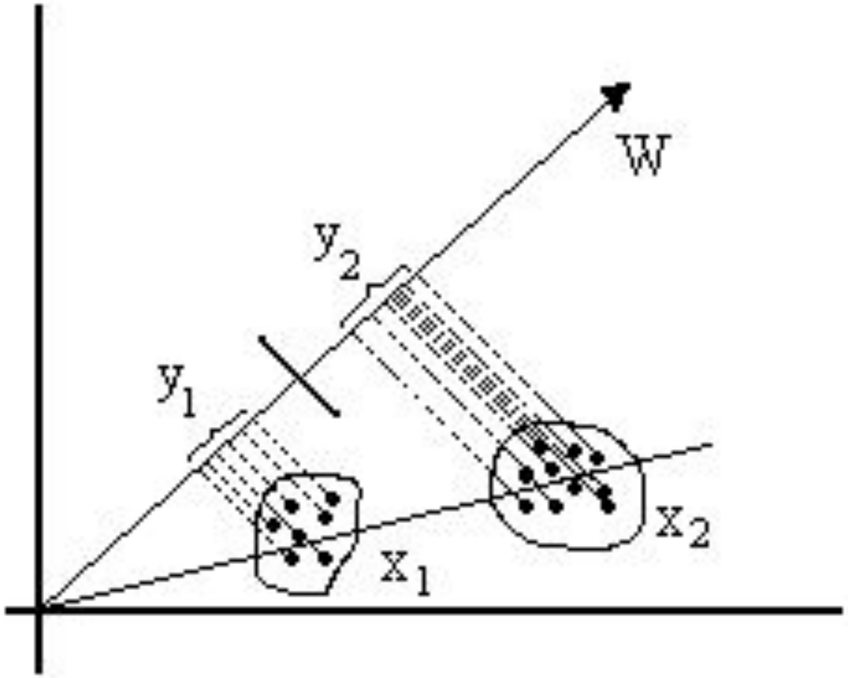
$$\overline{W} = S_W^{-1}(\overline{m}_1 - \overline{m}_2)\frac{C}{\lambda},$$

- Так как вектор $\overline{W}$ произвольной длины, положим $\frac{C}{\lambda}$ =1, тогда имеем :

$$\overline{W} = S_W^{-1}(\overline{m}_1 - \overline{m}_2).$$

- Соответственно линейный дискриминант Фишера получается в следующем виде:

$y = \overline{W}^T \overline{X} = X^T W = \overline{X} S_W^{-1}(\overline{m}_1 - \overline{m}_2)$ - это проекция вектора $\overline{X}$ на ось W



Мы должны выбрать некоторый порог решения η

Правило решения имеет вид $X^T S_W^{-1}(\bar{m}_1 - \bar{m}_2) = \eta$

- Матрица $S_W = S_1 + S_2 \sim N\,\Sigma$ пропорциональна суммарной матрице ковариации и соответственно для случая нормального распределения исходных данных дискриминант Фишера дает результат такой же, как и байесовская линейная дискриминантная функция:

$$X^T \Sigma^{-1}(M_1 - M_2) + C_1 \ldots \geq C_2$$