

# Линейный дискриминантный анализ для р > 1

Расширим LDA-классификатор на случай с несколькими предикторами. Для этого предположим, что вектор $X = (X\_{1}, X\_{2}, \cdots , X\_{p})$ извлечен из многомерного гауссова распределения со специфичным для каждого класса вектором математического ожидания и общей ковариационной матрицей:

$f(x)=\frac{1}{(2π)^{p/2}\sqrt{detΣ}}exp\left(-\frac{1}{2}(x-μ)^{T}Σ^{-1}(x-μ)\right)$.

При р > 1 LDA-классификатор предполагает, что наблюдения в *k*-м классе имеют многомерное гауссово распределение $N(μ\_{k},Σ)$, где $μ\_{k}$ — это специфичный для данного класса вектор математического ожидания, а $Σ$ — общая для всех *K* классов ковариационная матрица. Подстановка функции плотности вероятности для *k*-гo класса $f\_{k}(X = x)$ в $p\_{k}(x)$ и выполнение некоторых алгебраических преобразований показывает, что байесовский классификатор относит наблюдение $X = x$ к классу, у которого величина

$δ\_{k}(x)=x^{T}Σ^{-1}μ\_{k}-\frac{1}{2}μ\_{k}^{T}Σ^{-1}μ\_{k}+lnπ\_{k}$

максимальна. Это — векторно-матричная форма записи $δ\_{k}(x)$ из прошлой лекции.

Необходимо оценить неизвестные параметры $μ\_{1},\cdots ,μ\_{K}$, $π\_{k}, \cdots , π\_{K}$ и $Σ$; формулы для них похожи на соответствующие формулы для одномерного случая. Для классификации нового наблюдения X = х LDA подставляет эти оценки и относит это наблюдение к классу с наиболее высоким значением $\hat{δ\_{k}}(x)$. $δ\_{k}(x)$ является линейной функцией от *х*, т.е. решающее правило LDA определяется только линейной комбинацией элементов х. Это также является причиной использования прилагательного “линейный” в названии LDA.

LDA предполагает, что наблюдения в каждом классе происходят из многомерного гауссова распределения со специфичными для каждого класса векторами математических ожиданий и общей для всех *K* классов ковариационной матрицей. Квадратичный дискриминантный анализ (QDA) обеспечивает альтернативный подход. Подобно LDA, QDA-классификатор получают, делая предположение о том, что наблюдения в каждом классе имеют гауссово распределение, и подставляя оценки соответствующих параметров в уравнение теоремы Байеса для вычисления предсказаний. Однако, в отличие от LDA, QDA предполагает, что каждый класс имеет свою собственную ковариационную матрицу. Другими словами, этот метод предполагает, что наблюдение из *k*-го класса имеет форму $X \~ N(μ\_{k}, Σ\_{k})$, где $ Σ\_{k}$ — это ковариационная матрица *k*-гo класса. Исходя из этого предположения, байесовский классификатор относит наблюдение *X* = *х* к классу, для которого величина

$δ\_{k}(x)= -\frac{1}{2}(x-μ\_{k})^{T}Σ\_{k}^{-1}(x-μ\_{k})-\frac{1}{2}lndetΣ\_{k}+lnπ\_{k}$

максимальна. Таким образом, QDA-классификатор подразумевает подстановку оценок $Σ\_{k}$, $μ\_{k}$ и $π\_{k}$ в формулу и последующее отнесение наблюдения X = х к классу, для которого полученная величина является наибольшей. В данном случае член *х* представлен в виде квадратичной функции. Отсюда и происходит название QDA.