Многомерное нормальное распределение

Это обобщение нормального распределения для пространства большей размерности. Не путать со смесью нормальных распределений (GMM), это как путать задачу многомерной классификации (на вход подается вектор) и задачу мультиклассификации, когда много классов. Гауссианы используются для оценки плотности, а через нее решается задача классификации. Также гауссианы используются для генерации точек.

Вычисление плотности

$$f_{\mu,\Sigma}(x) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^k|\Sigma|}}$$

$\sum$ - матрица ковариации, на диагонали стоят дисперсии. Для одномерного случая $\sum$ заменяется на дисперсию.

μ - вектор смещения, матожидание по каждой координате

Генерация точки

- $\Sigma = A^{\mathrm{T}}A$
- $z_i \leftarrow \mathcal{N}(0,1)$
- $x \leftarrow zA + \mu$

1. Берем точку из простого нормального распределения, где все координаты независимы и $\sigma = 1$ по каждой координате. Оно задается кругом.
2. Круг поворачиваем (h_i) и растягиваем (λ_i)
3. Прибавляем вектор μ, чтобы сместить центр

$x \sum x^T$ задаёт эллипс, когда матрица $\sum$ положительно определена и симметрична.

h_i - собственные вектора, соответствуют осям нашего эллипса

λ_i - собственные числа, длины осей

Нормализация

Декорреляция

- $\mathbb{E}[X_j] = 0$ – центрирование матрицы
- Ковариационная матрица:

$$\Sigma(X) = \frac{1}{N}X^T X$$

Если дополнительно $\mathbb{D}[X_j] = 1$, то матрица корреляционная.

- Для безмодельного алгоритма делим на N, но если алгоритм модельный, то есть будет видить новые данные, то делим на $N-1$:

$$\Sigma(X) = \frac{1}{N-1} X^T X$$

- Декорреляция:

$$\hat{X} = X \times \Sigma^{-1/2}(X)$$

$\Sigma^{-1/2}$ - разложение Холецкого

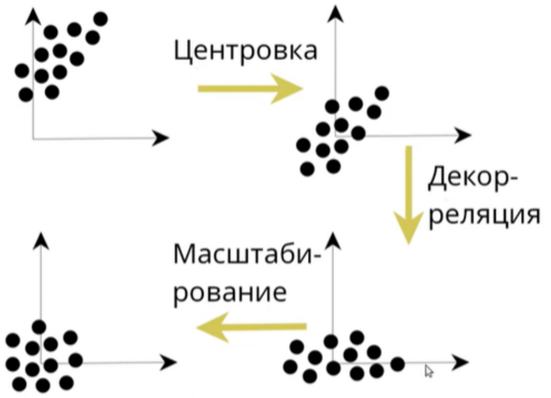
- $\Sigma(\hat{X}) = I$

▲

1. $\Sigma = \frac{X^T X}{N}$
2. $\hat{X} = X \times \Sigma^{-1/2}$
3. $\Sigma = \Sigma^{T/2}\Sigma^{1/2}$

$\frac{\hat{X}^T \hat{X}}{N} = \frac{(X\Sigma^{-1/2})^T (X\Sigma^{-1/2})}{N} = \frac{\Sigma^{-T/2} X^T X \Sigma^{-1/2}}{N} = \Sigma^{-T/2}\Sigma\Sigma^{-1/2} = \Sigma^{-T/2}\Sigma^{T/2}\Sigma^{1/2}\Sigma^{-1/2} = I$

■



Расстояние Махаланобиса

Расстояние Махаланобиса - Евклидово расстояние в декоррелированном пространстве.

$$\rho(a,b) = \sqrt{(a-b)S^{-1}(a-b)^T}$$

Декоррелируем объекты: $a \cdot S^{-1/2}, b \cdot S^{-1/2}$, подставим в расстояние Евклида:

$$\rho = \sqrt{(a-b)(a-b)^T} = \sqrt{(a-b)S^{-1/2}S^{-T/2}(a-b)^T} = \sqrt{(a-b)S^{-1}(a-b)^T}$$

Сравним 2 способа:

- 1. Подсчет расстояния Махаланобиса

$$X = [n, m]$$

$$\text{Предподсчеты: } nm^2(\text{матрица ковариации}) + m^3(\text{вычисление обратной})$$

$$\text{Расстояние: } m^2 - \ldots \cdot S^{-1} \cdot \ldots(\text{умножение с двух сторон})$$

- Сначала декорреляция, потом расстояние Евклида

$$\text{Предподсчеты: } nm^2(\text{матрица ковариации}) + m^3(\text{обратная}) +$$

$$+m^3(\text{разложение Холецкого}) + nm^2(\text{применение декорреляции})$$

$$\text{Расстояние: } m$$

Сингулярное разложение

Теорема: любая матрица F размера $n \times m$ может быть представлена в виде сингулярного разложения

$$F = VDU^\top.$$

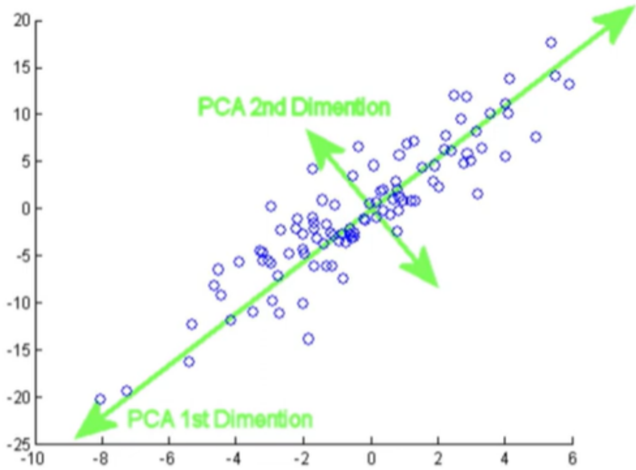
- $V = (v_1, \ldots, v_m)$ размера $n \times m$, являющаяся ортогональной: $V^\top V = I_m$, столбцы v_j — собственные вектора матрицы $FF^\top$
- $D = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m})$ размера $m \times m$, $\sqrt{\lambda_j}$ — **сингулярные** числа, квадратные корни собственных значений $F^\top F$
- $U = (u_1, \ldots, u_m)$ размера $m \times m$, являющаяся ортогональной: $U^\top U = I_m$, столбцы u_j — собственные вектора матрицы $F^\top F$

Представим какое-то скрытое пространство, в которое мы хотим проецировать данные.

- V представляет, как объекты соответствуют базисным векторам.
- D представляет важность каждого базисного вектора.
- U показывает, как признаки соответствуют базисным векторам.

Метод главных компонент PCA

- $X = VDU^{\mathrm{T}}$
- $\Sigma = \frac{1}{n}X^{\mathrm{T}}X = UD^{\mathrm{T}}V^{\mathrm{T}}VDU^{\mathrm{T}} = U\frac{D^2}{n}U^{\mathrm{T}}$
- $\Sigma^{\frac{1}{2}} = U^{\mathrm{T}}\frac{D}{n}$



Следовательно можно выбрать нужное число собственных векторов U_j с соответствующими собственными числами $\sqrt{\lambda_j}/n$.

Если изначально матрица была ЛЗ, то при вычислении $\Sigma^{-1/2}$ для декорреляции алгоритм сломается. Поэтому нужно выбросить какие-то собственные вектора, чтобы матрица стала ЛНЗ. В машинном обучении значения записываются с некоторой погрешностью, поэтому нельзя точно определить, какой вектор ЛЗ. Нужно как-то "мягко"принимать это решение.

Решение:

Если какое-то собственное число слишком маленькое, то значит собственный вектор является нормалью к какой-то плоскости, в которой лежат наши объекты. Так как объекты измерялись с погрешностью, мы не можем эту плоскость в явном виде увидеть. Поэтому мы выбираем некоторое количество собственных векторов, например, 3 с наибольшими собственными числами.