HW8A. 고유분해 응용 - decorrelation

- **Q1**. 다음 미분방정식의 해를 구하시오 (단, x(0) = 1, y(0) = 0)

$$\dot{x} = x + 3y$$

$$\dot{y} = 3x + y$$

HW8A. PCA 응용 - 부분공간과 차원감소

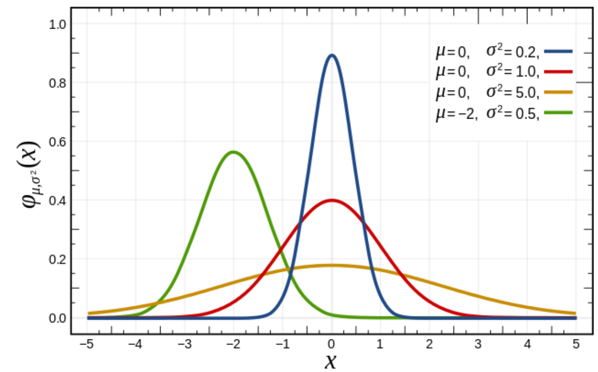
- **Q2.** 사과의 품종 A, B를 비교하기 위해 전년도에 수확한 A품종 1,000개체, B품종 500개체 대해 당도, 밀도, 색상, 수분함량을 측정한 데이터가 있다 (데이터 파일: data_a.txt, data_b.txt).
  - PCA를 이용하여 이들 1,500개 데이터의 분포를 가장 잘 설명할 수 있는 2개의 주성분 벡터(principal component) v1, v2를 구하시오.
  - 두 주성분 벡터 v1, v2에 의해 생성되는 부분공간 S를 생각하자. 원래의 1,500개 데이터를 부분공간 S에 투영한 좌표(coordinate)를 구하고 이들을 2D 평면에 가시화(visualization)하시오. 단, 구분이 용이하도록 A 데이터의 색상과 B 데이터의 색상을 달리하여 표시하시오.
  - 부분공간 S에서의 A 데이터의 분포와 B 데이터의 분포를 각각 2D Gaussian으로 모델링하시오.
  - 올해 새로 수확한 사과 2개체에 대해 당도, 밀도, 색상, 수분함량을 측정하였더니 test.txt와 같은 결과가 나왔다. 이 때, 새로 측정한 데이터를 부분공간에 투영한 후 앞서 구한 Gaussian 분포와의 'Mahalanobis 거리'를 계산하고 이를 이용하여 새로 수확한 사과 2개체의 품종을 각각 구분하시오.

참고) Gaussian 분포

- Gaussian 분포 = 정규분포(normal distribution)

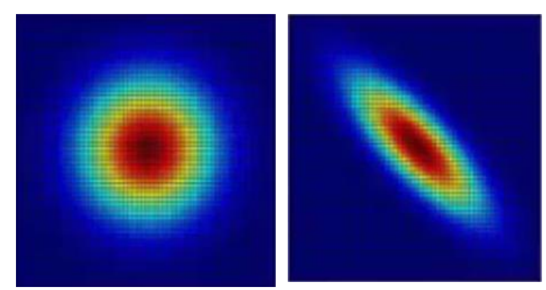
- 1-D 확률밀도함수

$$f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



- d-D 확률밀도함수

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$



Σ: covariance matrix
$|\Sigma|$: determinant