Aprendizado de Máquina e Reconhecimento de Padrões (UTFPR/CPGEI) - Lista de Exercícios 5

Tópicos: Extração de Características

1. Gere um conjunto de dados $\mathbf{X}_1$ bidimensional com 400 exemplos de duas classes distintas. Os primeiros 200 são da primeira classe, a qual é modelada por uma distribuição Gaussiana com média [-8, 8], e os outros 200 com média [8, 8]. A matriz de covariância de ambas as classes é S = [0.3, 1.5; 1.5, 9.0].
    1. Aplique PCA em $\mathbf{X}_1$ e calcule o percentual da variância total explicada por cada componente.
    2. Repita o processo para um conjunto $\mathbf{X}_2$. Altere as médias para [-1, 0] e [1, 0] e repita o processo.
    3. Compare os resultados.

2. Gere um conjunto de dados de 100 exemplos e dimensionalidade l=2000. Os vetores estão relacionados a um distribuição Gaussiana com média nula (l-dimensional) e uma matriz de covariância diagonal **S**, tendo todos os seus elementos diagonais iguais a 0.1, exceto S(1,1) e S(2,2), os quais são iguais a 10000. Aplique PCA e SVD e compare os resultados.

3. Gere um conjunto de dados tridimensional com 900 exemplos, os quais correspondem a duas classes – sendo os 100 primeiros com média nula e matriz de covariância $\mathbf{S}_1$ = [0.5 0 0; 0 0.5 0; 0 0 0.01]. Os demais estão organizados em 8 grupos de 100 exemplos. Cada grupo segue uma distribuição Gaussiana com a seguinte matriz de covariância: $\mathbf{S}_2$ = [1 0 0; 0 1 0; 0 0 0.01]. Sendo as médias:
    - [a 0 0];
    - [a/2 a/2 0];
    - [0 a 0];
    - [-a/2 a/2 0];
    - [-a 0 0];
    - [-a/2 –a/2 0];
    - [0 –a 0];
    - [a/2 –a/2 0].

Sendo a=6.

a) Plote os dados em 3-D e visualize-os sob vários ângulos para perceber a organização dos dados no espaço.
b) Aplique o discriminante de Fischer nos dados. Projete os dados em um subespaço gerado pelos autovetores que correspondem aos autovalores não-negativos do produto das matrizes $\mathbf{S}_w^{-1}\mathbf{S}_b$ (ver dica do final do exemplo visto em aula). Comente os resultados.

4. Gere dois conjuntos de dados **X** e $\mathbf{X}_{teste}$, sendo que cada um contém 200 exemplos tridimensionais. Em cada conjunto, os primeiros 100 exemplos são da classe 1, modelada através de uma distribuição uniforme em $[-0.5\ 0.5]^3$, enquanto os demais 100 exemplos são da classe -1

e estão localizados em torno de uma esfera de raio 2 centrada na origem. Os exemplos da segunda classe são gerados como segue. Selecione aleatoriamente um par de números, x(1) e x(2), de uma distribuição uniforme no intervalo [-2, 2] e cheque se $x^2(1)+x^2(2)$ é menor que r^2. Se esse não for o caso, selecione um par diferente. Caso contrário, gere dois pontos da esfera como $\left(x(1), x(2), \sqrt{r^2 - x^2(1) - x^2(2)} + \varepsilon_1\right)$ e $\left(x(1), x(2), \sqrt{r^2 - x^2(1) - x^2(2)} + \varepsilon_2\right)$, sendo ε_1 e ε_2 números aleatórios de uma distribuição uniforme no intervalo [-0.1 0.1]. Repita o processo 50 vezes para gerar 100 exemplos. Adicionalmente, gere os vetores de rótulos y e y_{teste} para os vetores X e X_{teste}. Em seguida, plote os dados.

1. Aplique a kernel PCA com um kernel exponencial e parâmetro 1, retendo apenas as duas primeiras componentes principais. Projete os dados no subespaço gerado pelas duas primeiras componentes principais, sendo **Y** o conjunto formado por essas projeções.
2. Projeto um classificador de mínimos quadrados baseado em **Y**.
3. Avalie a performance do classificador no conjunto de teste mapeado para o espaço 2-D gerado pela kernel PCA.
4. Repita o processo para o parâmetro do kernel igual a 0.6.

5. Gere uma espiral de Archimedes tridimensional como um pacote de 11 espirais de Archimedes bidimensionais idênticas, uma após a outra. Uma espiral bidimensional é descrita em coordenadas polares pela equação r=aθ, onde a é um parâmetro definido pelo usuário. No caso, os pontos em uma espiral 3-D são gerados como segue. Para cada θ, utilize os valores de θ_{init} até θ_{fin} com passo θ_{step} e calcule:

   - R = aθ
   - x = rcosθ
   - y = rsinθ

   Os 11 pontos são na forma (x, y, z), onde z = -1, -0.8, -0.6, ..., 0.8, 1 são pontos da espiral. Utilize a=0.1, θ_{init} = 0.5, θ_{fin} = 2.05*pi, θ_{step} = 0.2. Plote a espiral 3-D de tal forma que todas as espirais 2-D sejam plotadas com o mesmo símbolo e todos os grupos de 11 pontos na forma (x, y, z), onde x e y são fixados e z assume os valores -1, -0.8, -0.6, ..., 0.8, 1, são plotados com a mesma cor.

   a) Aplique kernel PCA para uma dimensão m=2 para diferentes parâmetros do kernel Gaussiano e plote os resultados.
   b) Repita o processo com PCA para os dois componentes principais.
   c) Compare e comente os resultados.

Exemplo da espiral 2-D e 3-D: