

Parte I - Análise de Componentes Principais

Uma das aplicações mais interessantes e surpreendentes de análise de componentes principais é no reconhecimento facial (*eigenfaces*). Nesta parte, você implementará este algoritmo para um pequeno conjunto de imagens de rostos.

- (a) Carregue os dados do arquivo *dado1.mat*. A variável X contém uma matriz em que cada linha corresponde a imagens de faces em tons de cinza, com tamanho original 32×32 , mas convertidas para um vetor de 1024 componentes. Mostre 100 imagens quaisquer desta base (lembre-se de reconverter o vetor linha para uma imagem). Implemente o algoritmo de PCA como visto em aula e rode sobre esta base de faces.
- (b) Os componentes PCA (autovetores) principais neste tipo de aplicação são o que se chama de *eigenfaces* e eles podem ser visualizados se forem reformatados para uma imagem 32×32 . Mostre em seu código as *eigenfaces* correspondentes aos 36 primeiros componentes principais. Comente no relatório sobre o que observa nestas imagens.
- (c) Projete a matriz X sobre os 100 componentes principais e reconstrua o dado original sobre esta projeção. Mostre as imagens originais de 100 faces ao lado das imagens reconstruídas. Comente no relatório sobre o que achou da reconstrução e implicações práticas deste processo.

Parte II - Sistemas de Recomendação

Nesta parte, você desenvolverá um sistema de recomendação de filmes usando filtragem colaborativa.

O arquivo *dado2.mat* contém notas de 1 a 5 dadas por usuários para filmes. A matriz Y armazena na linha i e coluna j a nota dada pelo usuário j para o filme i . Já em relação à matriz R , temos $R(i, j) = 1$ se o usuário j deu alguma nota para o filme i e 0 caso contrário.

- (a) Implemente o algoritmo de filtragem colaborativa. Este deverá aprender uma matriz X , que em cada linha contém o vetor de atributos $x^{(i)}$ do i -ésimo filme e uma matriz Θ , que em cada linha guarda o vetor de parâmetros $\theta^{(j)}$ para o j -ésimo usuário. Considere que tanto $x^{(i)}$ quanto $\theta^{(j)}$ possuem dimensão 100. Considere a função de custo sem regularização e use gradiente conjugado para minimizá-la (o mesmo do Projeto 2).
- (b) Com base nas notas preditas por seu algoritmo, liste os 10 filmes com notas médias mais altas, mostrando o nome e a nota média do respectivo filme. Para obter o nome do filme, use o arquivo *dado3.txt*, que contém o nome correspondente a cada linha na matriz Y .