Mini Relato a respeito da Análise das Componentes Principais de uma matriz contendo os vetores de Amostras:

- Objetivo: A Análise de componentes Principais é uma técnica que busca extrair informação de um montante de dados. Através desta análise é possível enxergar coisas que não são visualmente perceptíveis, a final os dados estão em um espaço vetorial de dimensão alta, muito menos de fácil inferência por análise nos dados. Buscar uma representação dos dados de baixa dimensão dos dados é importante pois nos habilita a visualização, além de diminuir a carga de cálculo que é feita nos dados crus para processar e extrair informações úteis. A pergunta que deve ser feita é: Qual/Quais as características mais importante do meu conjunto de pontos? É possível inferirmos com precisão? Em minha cabeça, antes de responder precisamos ter o mínimo de contato com os dados trabalhados que serão manipulados e, a partir de um 'tato' maior dos seus dados, é mais fácil estabelecer critérios que maximizarão, ou minimizarão, determinadas características que esperamos que os dados contenham.
- Componentes Principais: A análise PCA busca uma representação dos dados em dimensão alta em uma série de direções principais ortogonais cuja dimensão é menor que a original, ou seja, as projeções dos dados são descorrelacionadas (supondo que nossos dados originais possuam correlação entre si) idealmente, além de maximizar a variância nas direções principais.
- Parte que eu estudei, PCA através da Decomposição em Valor Singular da Matriz dos dados.
- Técnica: Suponha que tenhamos uma matriz $\boldsymbol{X}$, $\boldsymbol{N\mathrm{x}P}$, em que cada linha corresponde a uma amostra de seu conjunto de pontos de dimensão D. A decomposição em valores singulares de uma matriz é uma forma de decomposição matricial de tal modo ao ser aplicado conseguimos 2 bases ortonormais e uma matriz diagonal. Os vetores que compõem cada uma das bases são chamados direções singulares e cada elemento da matriz diagonal são os valores singulares. $\boldsymbol{X = U\Sigma V^t}$, isso é ótimo pois a decomposição em valor singular existe para qualquer Matrix, ou seja, é independente da quantidade de linhas e colunas, além de exigir independência linear entre as linhas ou colunas da Matriz aplicada. Porém, podemos fazer de algumas propriedades interessantes para resolver o problema da análise PCA ao aplicarmos o SVD em uma matriz quadrada.
- Para o nosso caso: $Cov(X) = \boldsymbol{X_c' X_c}$, em que $\boldsymbol{X_c = X - ones(n, 1).mean(X)}$, ou seja, é a matriz sem as médias de cada um dos vetores dos seus dados. Esse passo é importante para retirar qualquer tendência presente nos dados. Temos também uma outra matriz que pode ser construída $\boldsymbol{G_r(X) = X_c X_c'}$. As dimensões de cada uma das matrizes são, PxP e NxN, respectivamente, e cada uma delas condensam informações que podem ser interpretadas como distâncias: Uma em relação ao espaço de características dos dados, outra em relação aos próprios dados. Uma vez que podemos aplicar o SVD em $\boldsymbol{X}$ e, por consequência, em $\boldsymbol{X_c}$, podemos expandir as contas:
- $\boldsymbol{Cov(X) = X_c' X_c = (U\Sigma V^t)^t (U\Sigma V^t) = (V\Sigma U^t)(U\Sigma V^t) = (V\Sigma^2 V^t)}$
- $\boldsymbol{G_r(X) = X_c X_c' = \left(U\Sigma^2 U^t\right);}$
- Este resultado é muito legal porque nos mostra que ao podermos realizar a decomposição em valor singular dos dados, o cálculo das componentes principais em relação a seus próprios valores singulares pode ser feito utilizando a matriz de Gram, em vez da matriz de Covariância. As direções principais são dadas pelos vetores da matriz U, ou V, sendo que os autovalores, ou valores singulares, são dados por

$\boldsymbol{sqrt}(\boldsymbol{\Sigma^2})$. As componentes principais são facilmente encontradas através da operação $\boldsymbol{U\Sigma}$, porque os maiores valores singulares concentram a informação de maior variabilidade na direção correspondente ao seu autovetor associado. Uma vez que os autovetores a esquerda e a direita formam uma base ortonormal, a projeção dos dados em suas componentes principais fornecem uma projeção ortogonal na direção de maior variância obtida pela operação $\boldsymbol{U\Sigma\ ou\ V\Sigma}$.