



Data Science Academy

www.datascienceacademy.com.br

Matemática Para Machine Learning

E-book

Eigenvectors, Eigenvalues, PCA,
Covariance e Entropy

Parte 4



Entropia e Ganho de Informação

Na teoria da informação, o termo entropia refere-se a informações que não temos. As informações que não temos sobre um sistema, sua entropia, estão relacionadas à sua imprevisibilidade: o quanto isso pode nos surpreender.

Se você sabe que uma determinada moeda tem cabeças gravadas em ambos os lados, então lançar a moeda não lhe dá absolutamente nenhuma informação, porque ela será cara toda vez. Você não precisa virar para saber. Nós diríamos que a moeda de duas cabeças não contém informação, porque não tem como te surpreender.

Uma moeda equilibrada de dois lados contém um elemento de surpresa em cada jogada. E um dado de seis lados, pelo mesmo argumento, contém ainda mais surpresa a cada jogada, que poderia produzir qualquer um dos seis resultados com igual frequência. Ambos os objetos contêm informações no sentido técnico.

Agora vamos imaginar que o dado está viciado, ele mostra "três" em cinco dos seis testes e descobrimos que o jogo é fraudulento. De repente, a quantidade de surpresa produzida com cada jogada desse dado é bastante reduzida. Vemos uma tendência no comportamento do dado que nos dá maior capacidade de previsão.

Entender o dado viciado é análogo a encontrar um componente principal em um conjunto de dados. Você simplesmente identifica um padrão subjacente.

Essa transferência de informação, do que não sabemos sobre o sistema para o que sabemos, representa uma mudança na entropia. O insight diminui a entropia do sistema. Obter informações reduz a entropia. Isso é ganho de informação. E sim, esse tipo de entropia é subjetivo, pois depende do que sabemos sobre o sistema em questão.

Assim, cada componente principal que corta o gráfico de dispersão representa uma diminuição na entropia do sistema, em sua imprevisibilidade.

Acontece que explicar a forma dos dados um componente principal de cada vez, começando com o componente que é responsável pela maior variância, é semelhante ao andar dos dados em uma árvore de decisão. O primeiro componente do PCA, como a primeira divisão if-then-else em uma árvore de decisão formada adequadamente, ocorrerá ao longo da dimensão que mais reduz a imprevisibilidade.

Entropia é um tema fundamental ao construir modelos Random Forest, um dos principais algoritmos de Machine Learning e também em algumas arquiteturas de Deep Learning.



Referências:

<https://math.stackexchange.com/questions/24456/matrix-multiplication-interpreting-and-understanding-the-process/24469#24469>

<https://pdfs.semanticscholar.org/9dfa/3d30681788aac5077ede7b0ba2f7c4ac501e.pdf>

<https://news.ycombinator.com/item?id=10080415>

<https://skymind.ai/wiki/eigenvector>

<https://www.cs.cmu.edu/~mgormley/courses/10601-s17/slides/lecture18-pca.pdf>

<https://arxiv.org/pdf/1407.2904.pdf>

http://www.uta.fi/sis/mtt/mtts1-dimensionality_reduction/drv_lecture3_jan28update.pdf

<http://mathworld.wolfram.com/MatrixDiagonalization.html>