Entropia e árvores de decisão

Nos dois últimos posts vimos duas ideias importantes.

2026-03-18

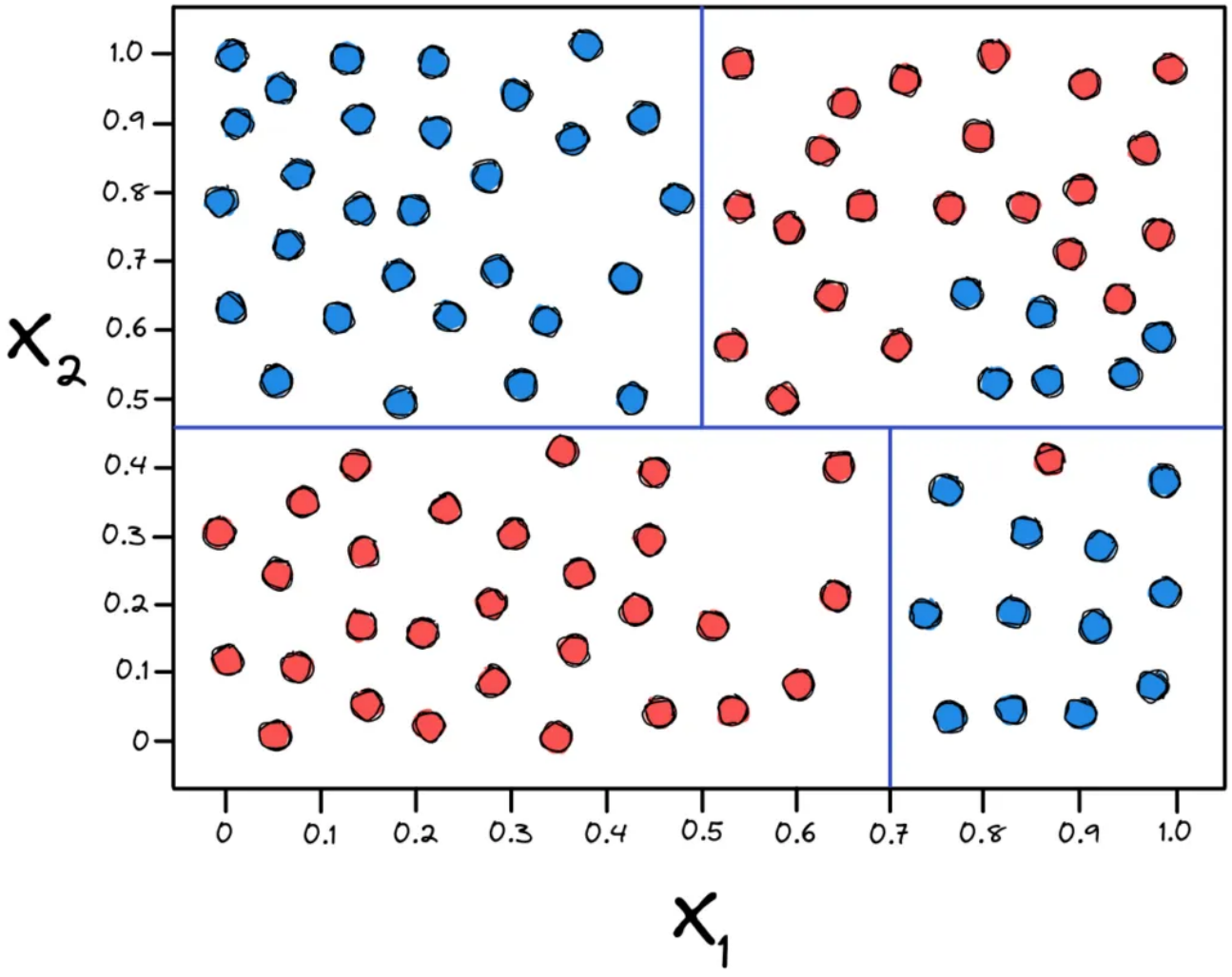


Figure 1: Capa do artigo

Nos dois últimos posts vimos duas ideias importantes.

Com Landauer, entendemos que informação tem custo físico e com Shannon, vimos como medir a incerteza nos dados.

Agora vamos dar o próximo passo:

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

como isso aparece quando um modelo de floresta tomam decisões.

Vem comigo!

No post anterior usamos o exemplo das bolas coloridas, agora vamos para outro cenário:

Imagine uma lista de pessoas com idade e se compraram um produto ou não, sem organizar, fica difícil entender o padrão.

Agora imagine separar assim:

- pessoas com menos de 30 anos
- pessoas com 30 anos ou mais

De repente, os grupos começam a fazer mais sentido porque você começa a enxergar padrões.

É exatamente isso que um modelo faz.

Ele cria divisões usando perguntas simples:

- “idade é maior que 30?”
- “valor é maior que 100?”
- “pertence a esse grupo?”

Cada pergunta separa melhor os dados.

Depois de algumas divisões, os grupos ficam mais organizados e alguns grupos passam a ter comportamentos bem parecidos.

Isso facilita prever o que pode acontecer com novos dados.

Aqui novamente entra Shannon pois organizar os dados significa reduzir a incerteza.

Quanto mais claros os grupos, mais previsível o comportamento, e modelo está reduzindo a entropia dos dados

Para chegar nesses grupos organizados, o computador executa operações.

Ele lê dados.

Compara valores.

Toma decisões.

Tudo isso envolve processamento físico, ou seja o princípio de Landauer. Existe energia sendo usada para organizar a informação.

No dia a dia isso aparece de forma direta Quando você filtra um dataset,quando cria uma regra ou ainda quando treina um modelo.

Além disso pra gente decidir como dividir os dados, o modelo precisa medir o quão “misturado” está um grupo.

Uma forma de fazer isso é usando a impureza de Gini.

Pense em um grupo com pessoas que compraram e não compraram um produto, se metade comprou e metade não, o grupo está bem misturado.

Se quase todo mundo tomou a mesma decisão, o grupo está mais organizado.

O Gini mede exatamente isso.

Ele indica o quão provável é errar ao tentar classificar alguém daquele grupo.

Quanto mais misturado o grupo, maior a chance de erro.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

Quanto mais homogêneo, mais fácil prever.

O modelo testa várias divisões e escolhe aquela que cria grupos mais consistentes.

Quando um modelo de floresta oma uma decisão, ele está reorganizando um conjunto de informações que antes estava misturado, e isso tudo é sobre medir incerteza que depois se traduz em puro conceitos de termodinâmica.

Lindo demais né?

Eu sou gamada nisso.

E ocê o que acha?

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.