

Mergulho nas IDEs Online para Aprendizado de Máquina (I)

Yuri Vacelh Zamulhak Zdebski

Descrição da atividade

Assistir ao video Jupyter Notebook Complete Beginner Guide - From Jupyter to Jupyterlab, Google Colab and Kaggle!, do canal Rob Mulla sobre as principais IDEs utilizadas para data science e machine learning.

Jupyter Notebook

Provavelmente a mais famosa das IDEs, o nome Jupyter Notebook é derivado das três principais linguagens que inicialmente suportava: Julia, Python e R. Essas linguagens são amplamente utilizadas para análise de grandes volumes de dados, cálculos estatísticos, desenvolvimento de modelos de machine learning, entre outros. A versão mais atual é a jupyter lab e ela conta com melhoras na aparência e em como navegar pelos notebooks (eu pessoalmente prefiro a versão mais simples).

O principal diferencial do Jupyter foi a introdução de células de código, o que torna a execução dinâmica. Isso significa que pequenas edições no programa não necessitam que todo o arquivo seja reinterpretado, mas sim apenas a célula modificada. Variáveis que já estão na memória são preservadas, possibilitando ajustes e execuções rápidas. Esse recurso melhora significativamente a natureza interativa da análise de dados. Outro diferencial das células é a possibilidade de escrever em Markdown, o que possibilita a criação de notebooks extremamente didáticos, pois além de ver a execução acontecer, sem a necessidade de um debugger, as anotações auxiliam a traduzir o que está acontecendo no código.

O modelo de Notebooks ficou tão popular que existem implementações em cloud (pois muito do do processamento realizado no contexto que os notebooks estão inseridos é pesado, e o hardware local simplesmente não está disponível), alguns exemplos memoráveis são:

- Google Colab: Plataforma da google em que salva os notebooks diretamente no drive, facilitando o compartilhamento e colaborações (dai que vem o colab)
- Kaggle: Uma plataforma online focada em hospedar e compartilhar datasets para os mais diversos propósitos (inclusive competições)

Os notebooks Jupyter também contam com diversos atalhos de teclado visando aumentar a produtividade. Esses atalhos podem ser divididos em várias categorias:

1. Execução de células:

- **Ctrl+Enter**: executa a célula atual
- **Shift+Enter**: executa a célula atual e move para a próxima

2. Navegação:

- **Esc**: entra no modo de comando
- **J e K**: navega entre células (no modo de comando)

- No Google Colab, é possível adicionar a navegabilidade do editor VIM, que também utiliza J e K como setas

3. Adição de células:

- `Esc + A`: adiciona uma nova célula acima da atual
- `Esc + B`: adiciona uma nova célula abaixo da atual

4. Alteração do tipo de célula:

- `Esc + M`: altera para célula de Markdown
- `Esc + Y`: altera para célula de código Python

Conclusões

Em suma, o Jupyter Notebook representa uma revolução na forma como programadores, cientistas de dados e pesquisadores interagem com código e dados. Sua capacidade de combinar células de código executável com texto formatado em Markdown cria um ambiente único para análise de dados, experimentação e documentação. A execução dinâmica das células de código permite uma abordagem interativa e iterativa para o desenvolvimento, facilitando ajustes rápidos e preservando o estado das variáveis em memória. Essa característica, juntamente com a possibilidade de criar narrativas explicativas usando Markdown, torna os Jupyter Notebooks uma ferramenta extremamente poderosa para:

- Análise exploratória de dados
- Prototipagem de modelos de machine learning
- Criação de relatórios interativos

Além disso, seu suporte a múltiplas linguagens, como Python, R e Julia, o torna versátil e adaptável a diversas necessidades computacionais. Em essência, o Jupyter Notebook não é apenas uma IDE, mas uma plataforma que redefine a maneira como pensamos sobre programação e análise de dados, tornando-as mais acessíveis, interativas e didáticas.