

Tirando dúvidas sobre o trabalho de aprendizagem:

Pessoal, no trabalho de vocês eu coloquei que “O conjunto de teste deve ser montado utilizando 30% do conjunto de treinamento selecionado aleatoriamente (a ferramenta permite a seleção randômica).”

Alguns conjuntos de treinamento já tem o conjunto de teste para ele com um número grande de exemplos. O problema é que o conjunto de teste existe somente para árvore de decisão e não para redes neurais. Por isso pedi que vocês escolhessem 30% do próprio conjunto de treinamento.

Para vocês traçarem a curva de aprendizagem será necessário simular o treinamento com tamanhos diferentes do conjunto de treinamento. Por exemplo, 3, 6, 9, 12 e 15 exemplos. O problema é que cada vez que precisarmos fazer o treinamento com um número diferente de exemplos, teremos que carregar novamente o conjunto de treino, e então, cada vez que se escolhe randomicamente o conjunto de teste, cada resultado será validado por um conjunto de teste diferente, o que não é nada coerente.

Para que todos sigam o mesmo princípio vamos estabelecer o seguinte:

1. Carregue pela primeira vez o conjunto de treinamento. Escolha randomicamente 30% do conjunto de treinamento para o conjunto de teste. **Anote quais foram os exemplos escolhidos para que os próximos treinamentos utilizem sempre o mesmo conjunto de teste** (para traçar a curva de aprendizagem);
2. Escolha os n primeiros exemplos, remova o restante e execute o treinamento;
3. Na próxima vez que o conjunto de treinamento for carregado, escolha manualmente os exemplos do conjunto de treinamento que você anotou no passo 1 para o conjunto de teste;
4. Volte para o passo 2.