

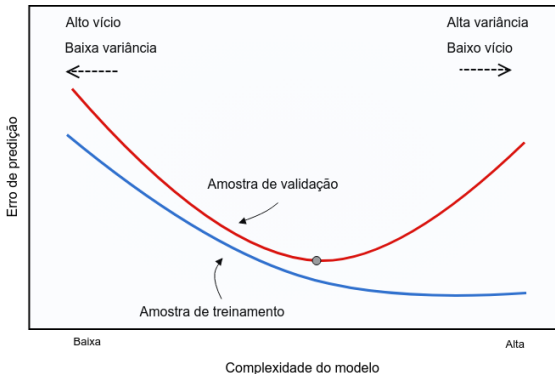
Como obter boas previsões

Prof. Eduardo Vargas Ferreira

Curso de Especialização em
Data Science & Big Data
Universidade Federal do Paraná

25 de agosto de 2018

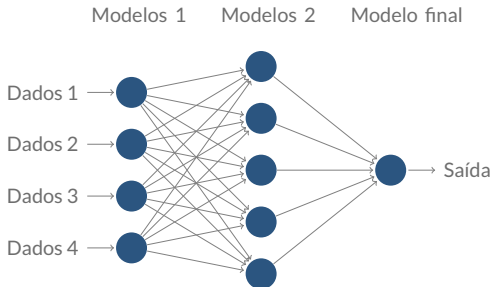
Bias-Variance Trade-Off



Como obter boas predições

1. **Entenda os dados:** explore as características, crie gráficos para entender a natureza das variáveis etc.;
2. **Decida sobre a validação cruzada:** uma boa estratégia de validação, garante resultados mais confiáveis, p ex.,
 - ▶ Se os dados mudam rápido com o tempo, ou são assimétricos etc., contemple isso nos exemplos de treinamento e validação.
3. **Feature Engineering:** tente aprimorar a acurácia do modelo, p. ex.
 - ▶ Tratar outlier, dados faltantes, criar interação, transformar dados contínuos para discretos etc. (pré-processamento);
4. **Combine modelos:** agrupe vários algoritmos, certificando-se que são correlacionados.

Combinando vários modelos

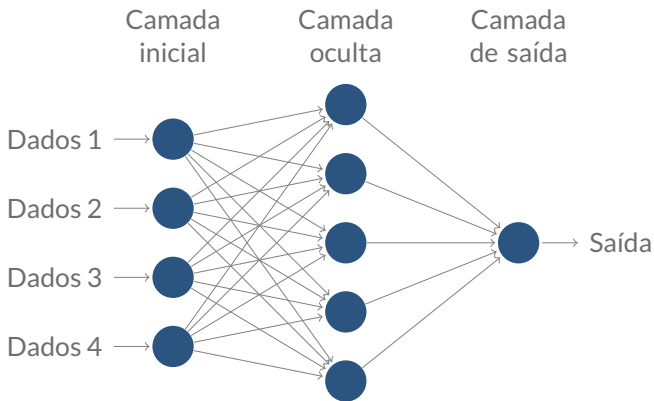


► As camadas de modelos envolvem técnicas do tipo:

- Regressão Linear;
- Regressão logística;
- KNN;
- Gradiente Boosting;
- Naive Bayes;
- Redes Neurais Artificiais;
- Árvores de decisão;
- Random Forests etc.

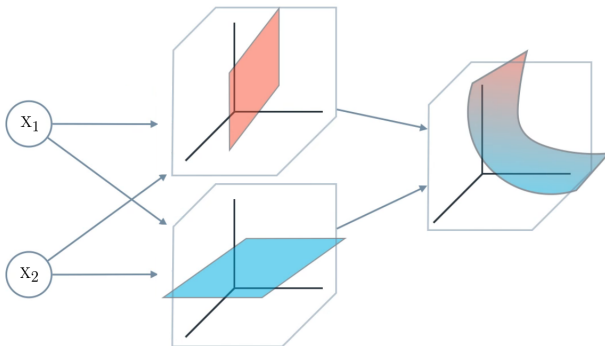
Redes Neurais Artificiais

O que são Redes Neurais Artificiais?



Exemplo

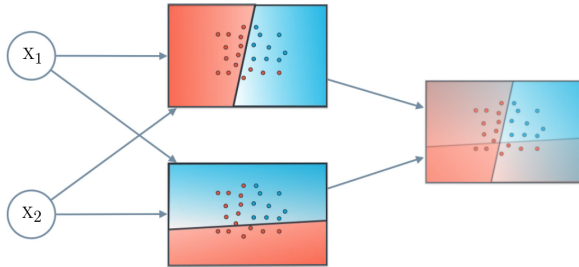
- Em problemas reais, a RNA combina separadores lineares para classificações mais complexas.



Fonte: Luis Serrano

Exemplo

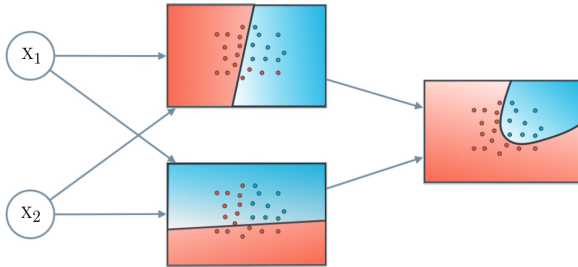
- Em problemas reais, a RNA combina separadores lineares para classificações mais complexas.



Fonte: Luis Serrano

Exemplo

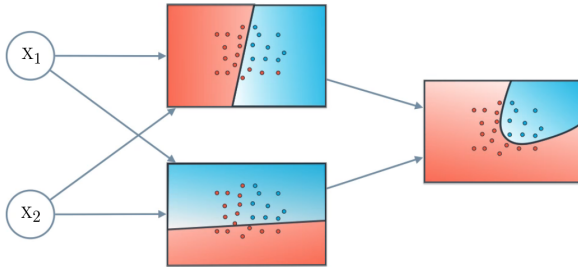
- Em problemas reais, a RNA combina separadores lineares para classificações mais complexas.



Fonte: Luis Serrano

Exemplo

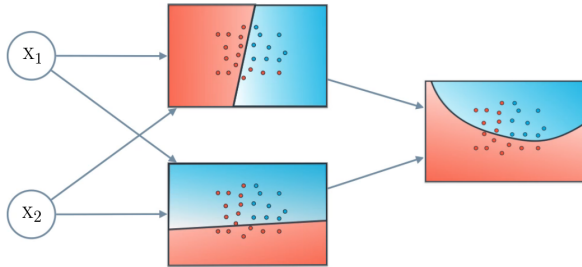
- Dependendo do peso atribuído a cada neurônio, obtemos diferentes regiões.



Fonte: Luis Serrano

Exemplo

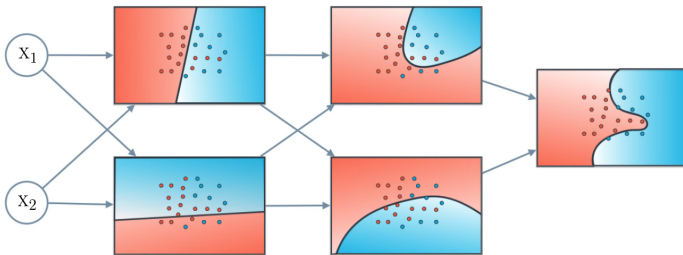
- Dependendo do peso atribuído a cada neurônio, obtemos diferentes regiões.



Fonte: Luis Serrano

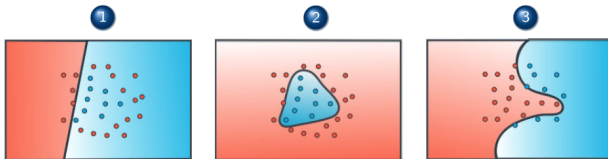
Quantas camadas devemos utilizar?

1. **Single layer:** capaz de posicionar um hiperplano no espaço das entradas;
2. **Two layers (one hidden layer):** capaz de descrever uma regra de decisão em somente uma região convexa do espaço;
3. **Three layers (two hidden layers):** a partir de três camadas, somos capazes de generalizar regiões arbitrárias do espaço.

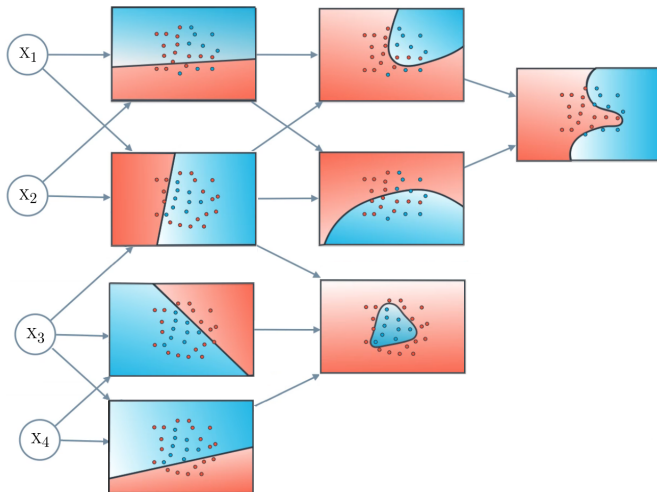


Quantas camadas devemos utilizar?

1. **Single layer:** capaz de posicionar um hiperplano no espaço das entradas;
2. **Two layers (one hidden layer):** capaz de descrever uma regra de decisão em somente uma região convexa do espaço;
3. **Three layers (two hidden layers):** a partir de três camadas, somos capazes de generalizar regiões arbitrárias do espaço.

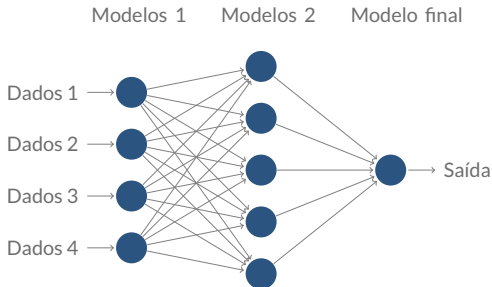


Exemplo



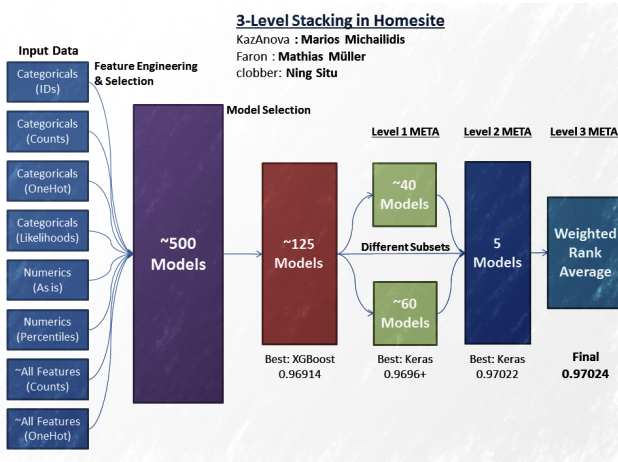
Fonte: Luis Serrano

Combinando vários modelos



- ▶ As camadas de modelos envolvem algoritmos do tipo:
 - ▶ Regressão Linear;
 - ▶ Regressão logística;
 - ▶ KNN;
 - ▶ Gradiente Boosting;
 - ▶ Naive Bayes;
 - ▶ Redes Neurais Artificiais;
 - ▶ Árvores de decisão;
 - ▶ Random Forests etc.

Exemplo



Linhas de comando

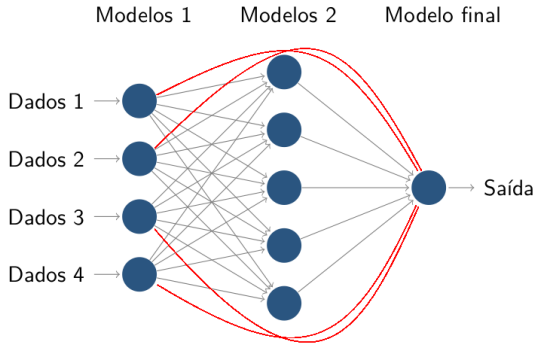
Linhas de comando

- Os algoritmos disponíveis requerem algumas especificações. Precisamos entendê-las, para obter resultados promissores, por exemplo:

Comando	Explicação
task	Pode ser regressão ou classificação.
metric	Métrica de saída na validação cruzada, para cada modelo-neurônio. Pode ser logloss, AUC, rmse etc.
stackdata	TRUE se saída do modelo na camada $k-1$ entrará também nas camadas $k+1$, $k+2$ etc.
bins	Parâmetro que permite que os classificadores sejam usados em problemas de regressão.
threads	Número de modelos a serem executados em paralelo.
folds	Número de <i>folds</i> no treinamento e teste.

Stackdata

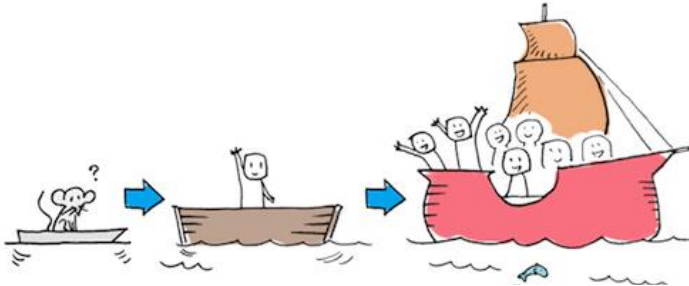
- TRUE se saída do modelo na camada $k-1$ entrará também nas camadas $k+1, k+2$ etc.



Desenvolver, testar e implementar

Desenvolver, testar e implementar

- ▶ É comum durante o treinamento dos modelos, obtermos determinado resultado, e na prática, o resultado ser outro;



- ▶ Para aumentar a convicção de que os modelos farão um bom trabalho, devemos ser capazes de treiná-los em condições próximas das reais;