Comitês de Máquinas

Lucas de Oliveira Santos
lucas.santos@lapisco.ifce.edu.br

Assuntos abordados

- ⇨ Introdução
- ⇨ Comitês de Máquinas (Ensembles)
- ⇨ Comitês Estáticos
- ⇨ Comitês Dinâmicos

Introdução

- Normalmente tomamos decisões importantes com o auxílio de pessoas que julgamos serem capazes de nos ajudar. Isso pode ser entendido como um “conselho” de pessoas que, embora pensem diferentes uns dos outros, nos ajudam a chegar em um senso comum.
- Embora pessoas e máquinas sejam diferentes, podemos estender a ideia de um “conselho” para as máquinas também. Dessa forma, diversos computadores poderiam solucionar um mesmo problema, apresentando resultados diferentes uns dos outros. Isso posto, damos o nome de comitê de máquinas à combinação de vários computadores, simulando resoluções distintas para uma mesma problemática.

Comitês de Máquinas

- Um ***comitê de máquinas*** baseia-se na ***combinação*** de diferentes ***estruturas de decisão***. Existem dois tipos de comitês: ***estáticos*** e ***dinâmicos***;
- ***Comitês estáticos***: não dependem das informações iniciais (dados de entrada) para gerar combinações de respostas;
- ***Comitês dinâmicos***: dependem das informações iniciais para integrar a resposta dos especialistas;

Comitês de Máquinas

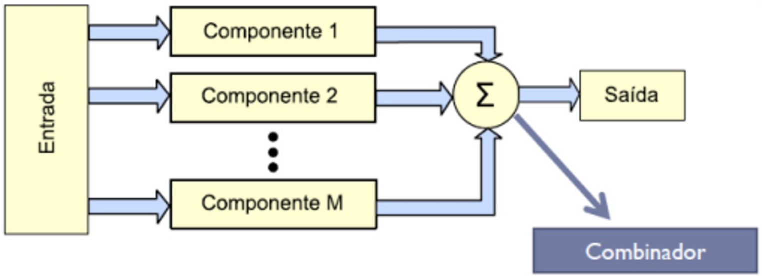

Figura 1. Comitê Estático (de [Coelho et al., 2016]).

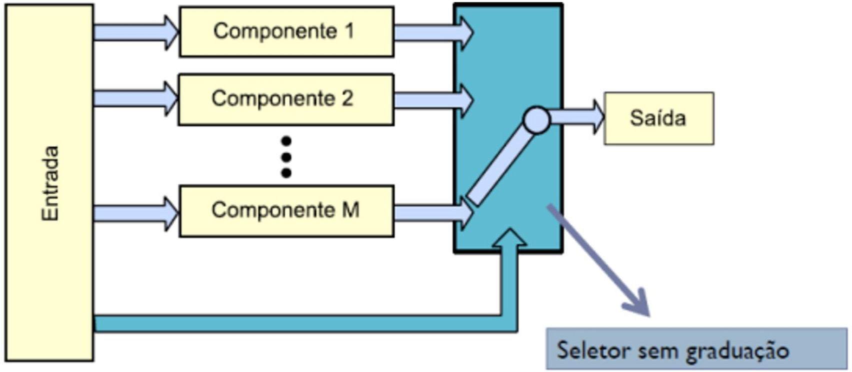

Figura 2. Comitê Dinâmico (de [Coelho et al., 2016]).

Comitês Estáticos

- Vamos considerar um problema de regressão em que se deseje aproximar uma função ideal ***f(x)*** a partir dos dados disponíveis. Um comitê com ***M*** máquinas possuindo modelos de regressão, nos forneceria as respostas estimadas $\boldsymbol{y_1(x)}$, $\boldsymbol{y_2(x)}$, $\boldsymbol{y_3(x)}$, …, $\boldsymbol{y_M(x)}$;

- Vamos considerar que a saída final do comitê é dada pela média aritmética das estimativas de respostas obtidas. Podemos considerar a saída final como sendo:

$$y_C(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} y_m(\mathbf{x})$$

- Dessa forma, a saída do ***m-ésimo*** modelo é dada por: $y_m(\mathbf{x}) = h(\mathbf{x}) + e_m(\mathbf{x})$; onde $\boldsymbol{e_m(x)}$ é o erro associado pelo ao ***m-ésimo*** modelo. Esse erro quadrático pode ser expresso por:

$$\mathrm{EQM}_m = E[{e_m}^2(\mathbf{x})] = E\left[\left(y_m(\mathbf{x}) - h(\mathbf{x})\right)^2\right]$$

Comitês Estáticos

- O erro médio quadrático individual é dado por: $\mathrm{EQM}_{\text{médio}} = \frac{1}{M}\sum_{m=1}^{M} E[{e_m}^2(\mathbf{x})] = \frac{1}{M}\sum_{m=1}^{M} EQM_m$;
- O erro médio associado ao comitê é dado por: $\mathrm{EQM}_C = E\left[\left(\frac{1}{M}\sum_{m=1}^{M} y_m(\mathbf{x}) - h(\mathbf{x})\right)^2\right] = E\left[\left(\frac{1}{M}\sum_{m=1}^{M} e_m(\mathbf{x})\right)^2\right]$
- ***Vamos supor*** que os erros ***e***$\boldsymbol{e_m(x)}$ possuem média zero e que eles não possuem correlação. Dessa forma, podemos inferir que: $\mathrm{EQM}_C = \frac{1}{M}\mathrm{EQM}_{\text{médio}}$;
- Podemos inferir que o ***comitê*** possui um ***erro quadrático médio*** menor do que a ***média dos erros individuais*** (***EQM***$\boldsymbol{EQM_{médio}}$), ou seja, a combinação de máquinas é proveitosa;
- Note que, supomos uma descorrelação entre os erros médios de cada componente do comitê. Isso é necessário para inserir diversidade nos dados para que cada modelo não seja aplicado sempre aos mesmos dados.

Comitês Estáticos

- Uma alternativa viável para possibilitar a descorrelação entre os erros é a utilização de um procedimento chamado ***bootstrap aggregation*** ou ***bagging***;
- Grosso modo, a técnica de ***bagging*** consiste em, dado um conjunto de dados com ***N*** amostras, gerar ***M*** (***número de comitês***) novos dados com a possibilidade de ***reposição de*** (dados repetidos). Levando-se em consideração dados repetidos, cada modelo treina com informações iniciais diferentes uns dos outros;
- Outra alternativa é a utilização do método ***boosting***. Essa técnica adota um esquema de ***treinamento sequencial***, ou seja, as máquinas realizam um treinamento uma após a outra. A etapa de treinamento se baseia em um ***dataset*** em que os dados são ponderados de acordo com o desempenho das máquinas anteriores;

Comitês Estáticos

- ⇨ Vamos exemplificar o método ***boosting*** utilizando um classificador qualquer. Nesse sentido, os dados classificados erroneamente pelos classificadores anteriores tem um peso maior no treinamento do modelo seguinte;
- ⇨ Imagine um conjunto com ***N*** dados. Associamos a esses dados pesos da forma W_n, $n = 1, 2, \ldots, N$. Considerando que cada peso corresponde à tarefa de projetar a ***m-ésima*** máquina, chegamos a seguinte notação: $w_n^{(m)},\ n = 1, \ldots, N$, com $m = 1, 2, \ldots, M$;

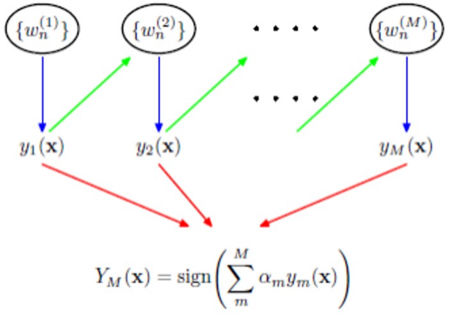

Figura 3. Estratégia de *Boosting* (de [Bishop, 2006]).

Comitês Estáticos - Exemplos

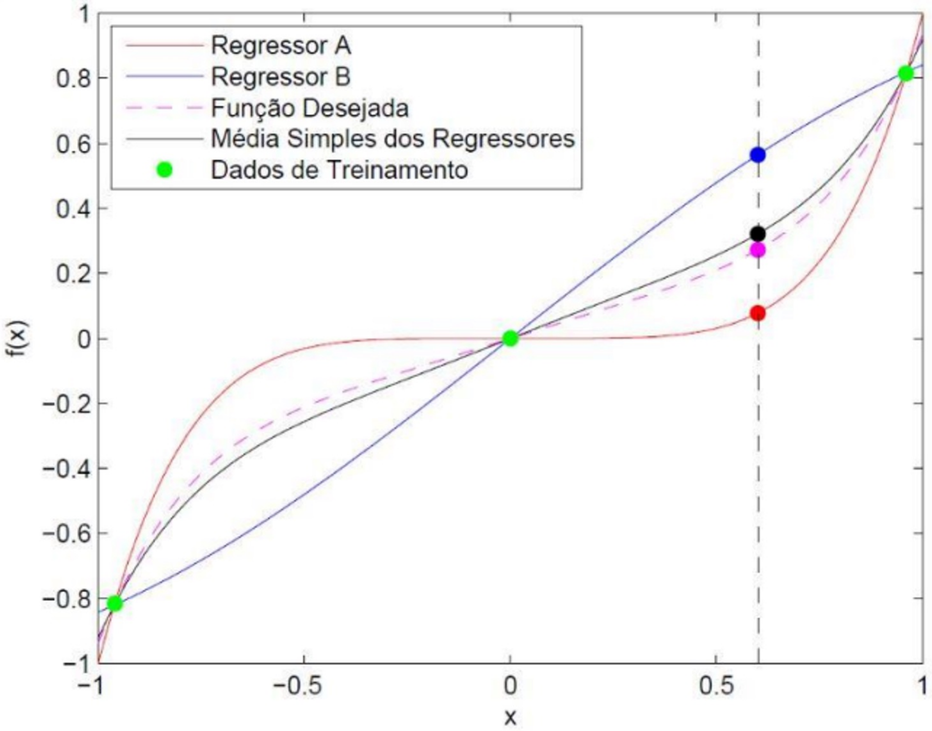

Figura. Aproximação de funções através da combinação de modelos por média simples.

Comitês Estáticos - Exemplos

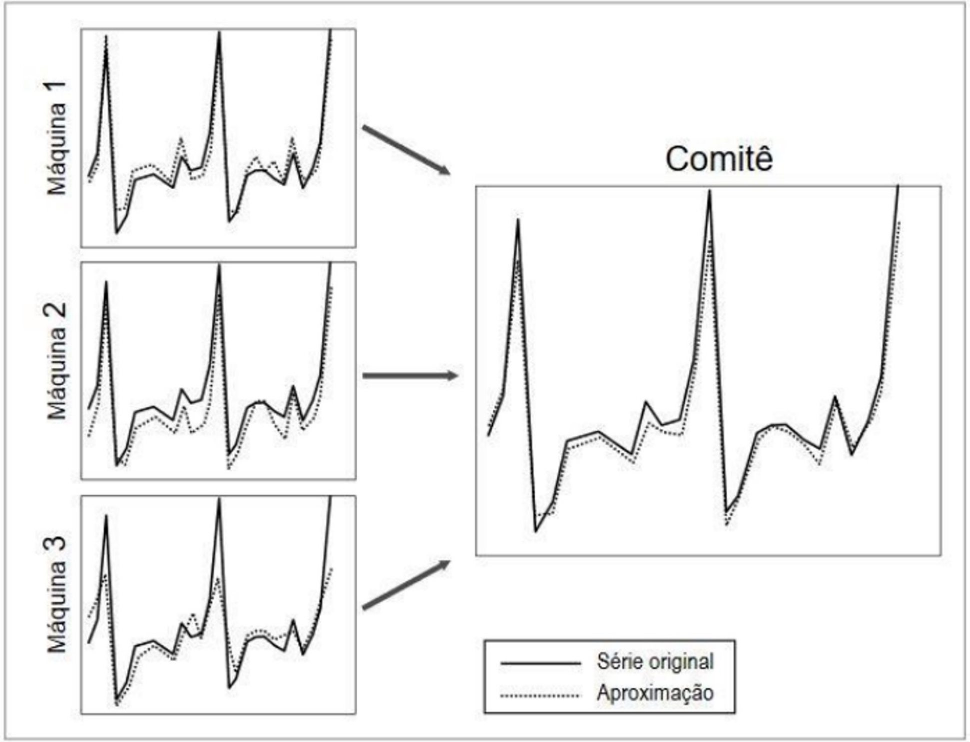

Figura. Combinação por média simples de modelos de previsão de séries temporais.

Comitês Estáticos - Exemplos

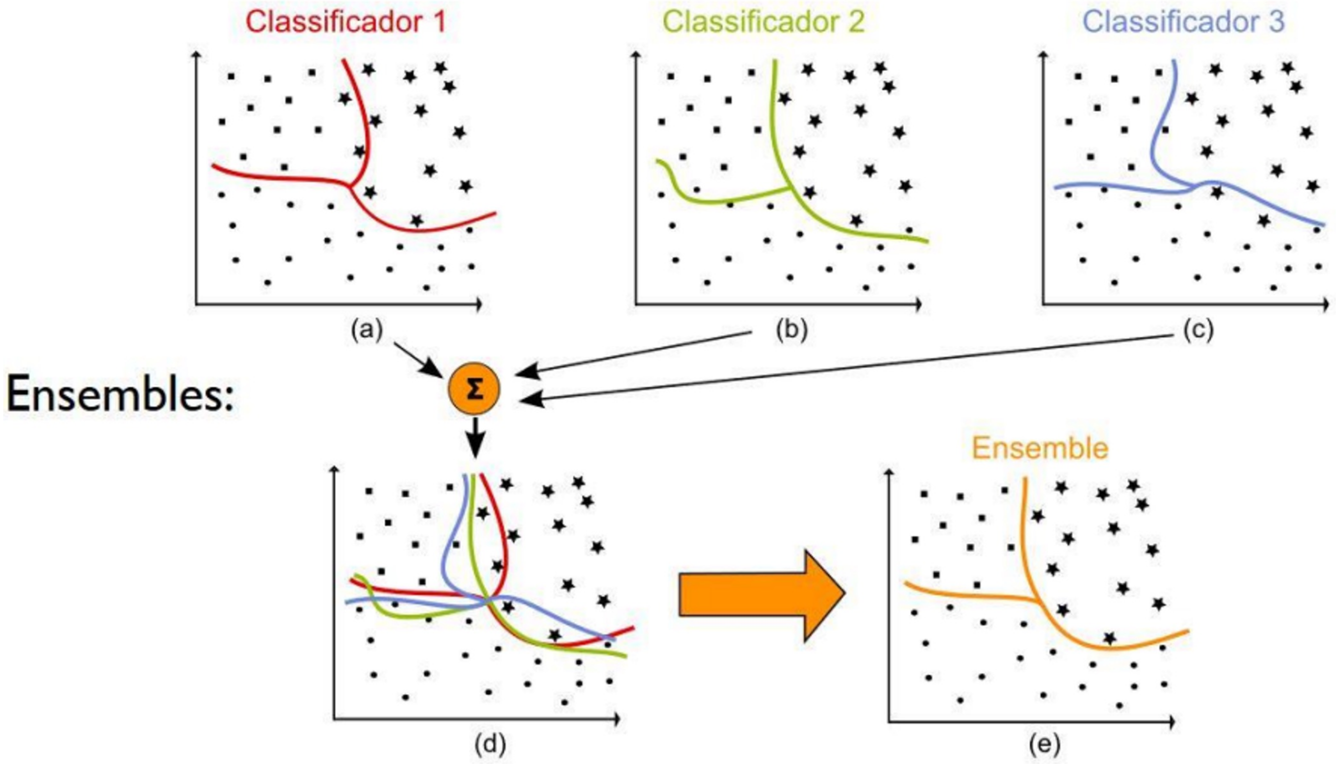

Figura. *Ensemble* de classificadores através de voto majoritário.

Comitês Estáticos

- As abordagem ***bagging*** e ***boosting*** geram diversidade levando em consideração os dados disponíveis. Existem outras formas de gerar diversidade no contexto de ***comitês de máquinas***.
- Levando-se em consideração o treinamento de ***redes neurais*** através da abordagem ***backpropagation***, valores diferentes podem ser utilizados para inicializar os ***pesos*** da rede. Dessa forma, obtêm-se modelos diferentes uns dos outros. Em suma, cada modelo pode apresentar ***condições iniciais diferentes*** no processo iterativo de treinamento.
- Pode-se variar o ***conjunto de arquiteturas*** usadas. Isso é possível pois os comitês podem ser ***homogêneos*** ou ***heterogêneos***.
- Do ponto de vista da ***homogeneidade***, pode-se usar estruturas idênticas mas com números distintos de unidades de processamento (MLP com diferentes quantidades de neurônios nas camadas intermediárias).
- Do ponto de vista da ***heterogeneidade***, diferentes modelos podem ser usados, como redes MLP e RBF por exemplo.

Comitês Dinâmicos

- ⇨ No caso de comitês dinâmicos, conhecido como ***mistura de especialistas***, tem-se uma abordagem diferente. Dessa forma, o problema é ***resolvido por partes*** por cada um dos componentes;
- ⇨ Cada modelo não resolve o problema como um todo, mas sim o ***conjunto de componentes*** é responsável por fazer isso;
- ⇨ É necessário projetar cada um dos componentes e também o estágio de controle conhecido como ***rede gating***. Essa rede indica, a cada entrada, o papel que os componentes do comitê devem desempenhar;
- ⇨ A ***rede gating*** pode ser implementada por uma arquitetura neural (linear ou não-linear) com uma camada de saída que utiliza a função de ativação softmax. Esse processo retorna os pesos g_m, $m = 1, 2, \ldots, M$ que serão tratados como os parâmetros de uma combinação linear das saídas dos especialistas (y_1, y_2, …, y_m). A expressão final é dada por: $$y_C = \sum_{m=1}^{M} g_m y_m$$

Comitês Dinâmicos

Mistura de Especialistas

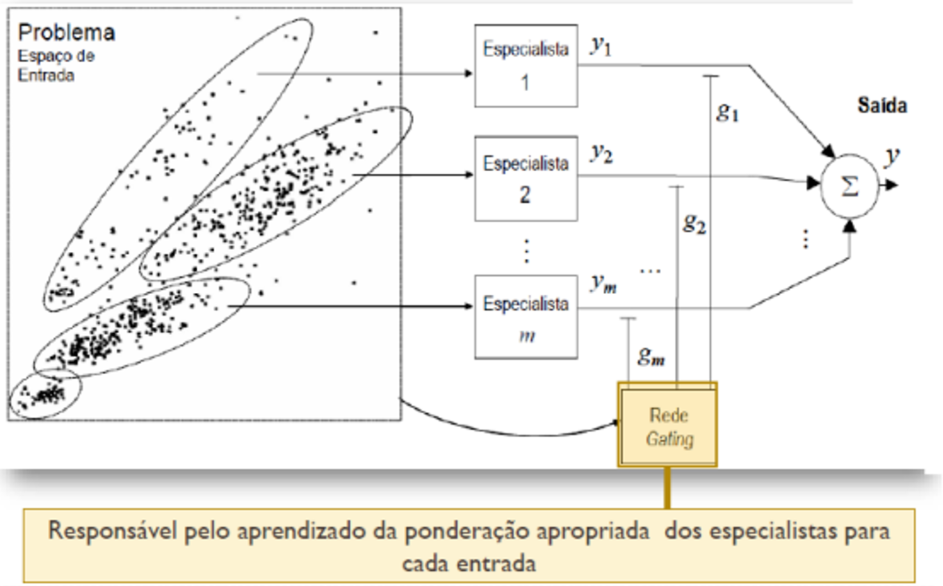

Figura 4. Esquema de Mistura de Especialistas (de [Coelho et al., 2016]).

Referências

- [1] A. Rocha Neto, “Sinpatco ii: Novas estratégias de aprendizado de máquinas para classificação de patologias da coluna vertebral,” Ph.D. dissertation, Tese (Doutorado)–Departamento de Engenharia de Teleinformática, Universidade . . . , 2011.
- [2] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain.” Psychological review, vol. 65, no. 6, p. 386, 1958.
- [3] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” The bulletin of mathematical biophysics, vol. 5, pp. 115–133, 1943.
- [4] M. Minsky and S. A. Papert, Perceptrons, reissue of the 1988 expanded edition with a new foreword by Leon Bottou: an introduction to computational geometry. MIT press, 2017.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.

Obrigado pela atenção!

Dúvidas?

Lucas de Oliveira Santos
lucas.santos@lapisco.ifce.edu.br