

-
-
-
-
-

Máquinas de Comitê

Aluizio Fausto Ribeiro Araújo
Universidade Federal de Pernambuco
Centro de Informática



-
-
-
-
-

Conteúdo

1. Introdução
2. Máquinas Estáticas de Comitê
 - 2.1. Votação
 - 2.2. Média do Conjunto
 - 2.3. Impulso (*Boosting*)
3. Máquinas Dinâmicas de Comitê
 - 3.1. Mistura de Especialistas
 - 3.2. Mistura Hierárquica de Especialistas

Introdução

- Ao escolher um modelo de rede neural (RN) para resolver um problema (e.g., uma MLP-BP), o projetista pode descartar outras redes que podem apresentar desempenho levemente inferior àquela escolhida.
- A avaliação da capacidade de generalização pode ser imprecisa por depender dos dados escolhidos.
- Uma possível solução é combinar as saídas de diferentes RNs buscando usufruir do aprendizado de todas. Este conhecimento de cada RN pode apresentar diferenças importantes.
- O desempenho do conjunto de RNs pode ser melhor que de uma única RN.

Introdução

- Uma máquina de comitê é um tipo de RN que emprega a estratégia dividir e conquistar na qual as respostas de RNs múltiplas (tidas como especialistas) são combinadas em uma resposta única. Esta resposta combinada deve ter desempenho superior a qualquer um de seus componentes.
 - Idéia de uma máquina de comitê remonta a Nilsson (1965).
 - A estrutura da rede consistia de uma camada de perceptrons elementares, seguida de um perceptron de votação na segunda camada.
- Dividir para conquistar: Uma tarefa computacional complexa pode ser resolvida dividindo-a em sub-tarefas mais simples.
 - Tais sub-tarefas são distribuídas entre unidades denominadas especialistas, que dividem sua atuação entre si para o espaço de entrada.

Introdução

- “Não há almoço grátis (*free lunch*): Não existe um algoritmo que induza o aprendiz mais acurado em qualquer domínio para todo o tempo.”
 - <http://www.no-free-lunch.org/>.
- Aprendizes distintos podem empregar diferentes:
 - Algoritmos (e.g., baseados em hipóteses); parâmetros (e.g., número de unidades escondidas); representações (e.g., características e fontes de informação); e conjuntos de treinamentos (e.g., combinações de dados).
- Abordagem com diferentes especialistas permite:
 - Explorar informações complementares; redundância e simultaneidade de avaliação.
 - Expressar preferência ou tendência dos especialistas considerando o custo computacional do consenso.

Introdução

As Máquinas de Comitê são classificadas em duas categorias:

- Estruturas Estáticas: As respostas de diferentes especialistas são combinadas por mecanismo que não envolve o sinal de entrada.

- Média do conjunto (*ensemble averaging*): Saídas de especialistas diferentes são combinadas linearmente para produzir a saída.

- Impulso (*boosting*): Aprendizagem por um algoritmo fraco (pouco mais de 50% de acertos) é convertida em um algoritmo com alta precisão.

- Estruturas Dinâmicas: O sinal de entrada faz parte do mecanismo que integra as saídas dos especialistas.

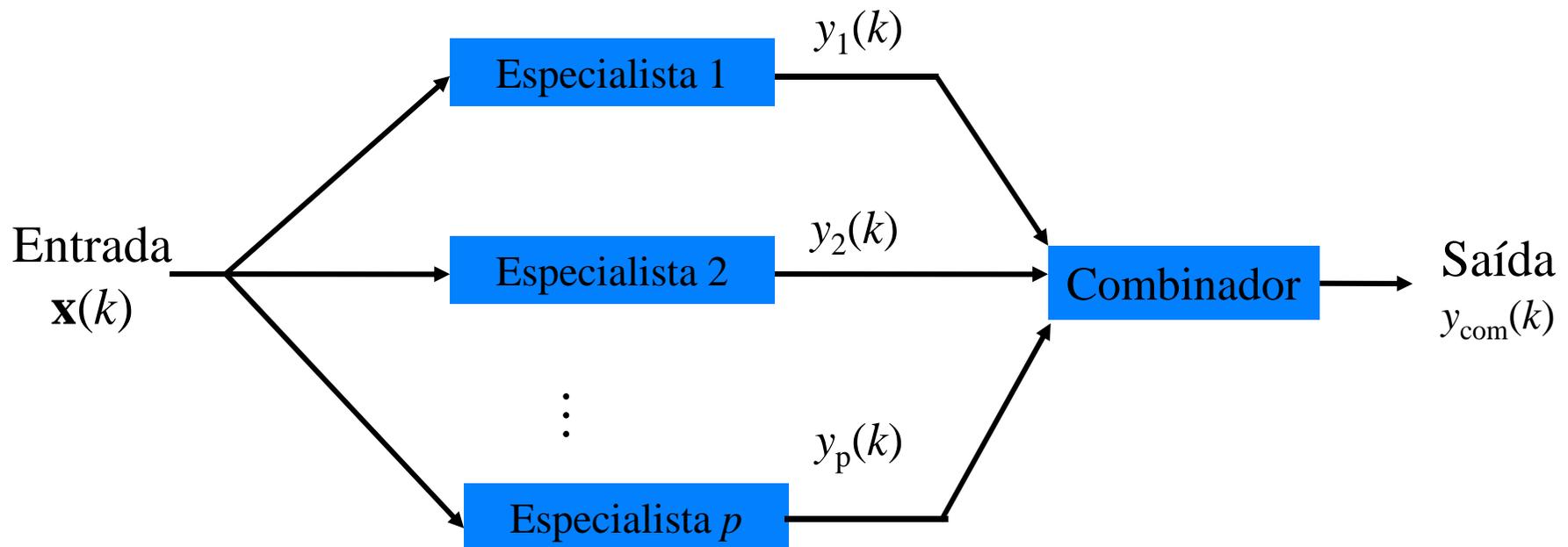
- Mistura de especialistas: Saídas dos diferentes especialistas são não-linearmente combinados por meio de uma rede única.

- Mistura hierárquica de especialistas: Saídas dos diferentes especialistas são não-linearmente combinados por meio de várias redes organizadas de

forma hierárquica.



Máquinas de Comitê Estáticas



Máquinas de Comitê Estáticas

- Redes neurais diferentes (especialistas) são treinadas de modo a compartilhar uma entrada comum e suas saídas individuais são combinadas de alguma forma para se produzir uma saída única.
- Motivações:
 - Grande número de parâmetros ajustáveis.
 - Risco de sobre-treinamento (*overfitting*) por causa do número.
- É esperado que:
 - Os especialistas convirjam para diferentes ótimos locais na superfície de erro.
 - Desempenho final melhor como consequência da combinação das saídas.

Máquinas de Comitê Estáticas

- Realiza uma média ponderada das saídas.
 - Para pesos iguais, classificações inadequadas podem prevalecer.
 - Análogo a votação a qual é usada para classificação.
- Pode gerar problemas na classificação pois se faz distribuição heterogênea do conhecimento especialista.
 - Um tipo de resposta particular não alcançada por um especialista com bom desempenho pode não ser atingida pela média se um especialista com baixo desempenho pode resolver a questão.

Máquinas de Comitê Estáticas: Tipos de Estruturas

- Votação
- Média do Conjunto (*Ensemble Averaging*).
 - Ensacando (*Bagging*)
- Impulso (*Boosting*)
 - Por filtragem (*Boosting by filtering*)
 - Por Sub-amostragem (*AdaBoosting*)
 - Por ponderação

Máquinas de Comitê Estáticas: Votação

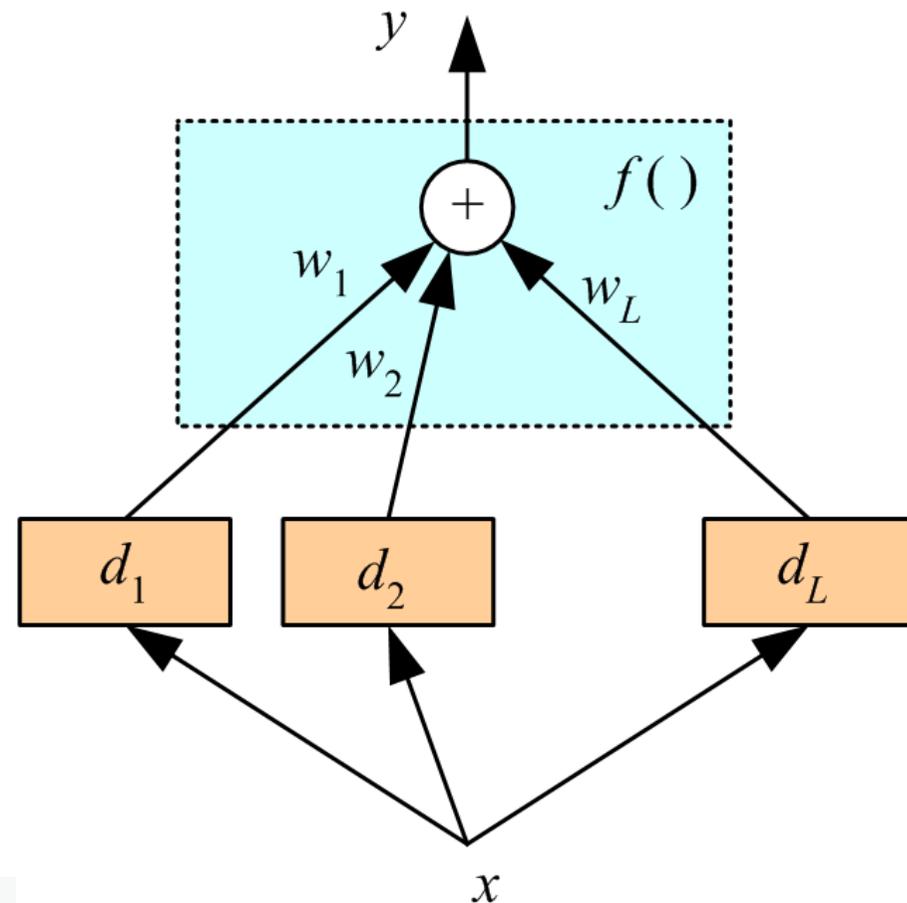
- Regressão

$$y = \sum_{j=1}^L w_j d_j$$

$$w_j \geq 0 \text{ and } \sum_{j=1}^L w_j = 1$$

- Classificação

$$y_i = \sum_{j=1}^L w_j d_{ji}$$



Máquinas de Comitê Estáticas: Votação

$$y = \sum_{j=1}^L w_j d_j$$

$$w_j \geq 0 \text{ and } \sum_{j=1}^L w_j = 1$$

$$y_i = \sum_{j=1}^L w_j d_{ji}$$

- Na regressão tem-se que $w_j = 1/L$:
 - Votação por maioria: A saída mais votada vence, i.e., cada saída de especialista afeta igualmente a saída global.
- Na classificação, cada peso, w_j , é proporcional à taxa de erro do classificador. Este é aprendido com respeito a um conjunto de validação.

Máquinas de Comitê Estáticas: Média do Conjunto

- Seja \mathbf{x} vetor de entrada não visto e \mathbf{d} a resposta desejada correspondente (representando uma classe, uma saída desejada), respectivamente realizações do vetor aleatório X e da variável aleatória D .
- Considere $F(\mathbf{x})$ a função de entrada-saída construída pela rede e baseado-se no dilema bias/variação.
 - Pode-se decompor o erro médio quadrático entre $F(\mathbf{x})$ e o valor esperado condicional $E[D|X=\mathbf{x}]$ nas componentes de bias e variância.

Máquinas de Comitê Estáticas: Média do Conjunto

- Combina-se múltiplos especialistas com bias e variância iguais:
 - O bias da função média de conjunto $F_1(\mathbf{x})$, relativo à máquina de comitê, é igual ao da função $F(\mathbf{x})$ relativo a uma única rede neural.
 - A variância da função média de conjunto $F_1(\mathbf{x})$ é menor que aquela da função $F(\mathbf{x})$.
 - Estratégia de treinamento para reduzir o erro global devido a condições iniciais variáveis: os especialistas são propositadamente sobretreinados.
- Bishop (pp.365-366) mostrou que a máquina de comitê por média de conjunto terá um erro esperado menor que a média dos erros de cada especialista individualmente.

Máquinas de Comitê Estáticas: Média do Conjunto - Exemplo

- Sejam duas classes:
 - $C_1: N(\mu = [0,0], \sigma^2 = 1)$
 - $C_2: N(\mu = [2,0], \sigma^2 = 4)$
- Critério de Bayes para fronteira de decisão ótima: $P(C_1|\mathbf{x}) > P(C_2|\mathbf{x})$

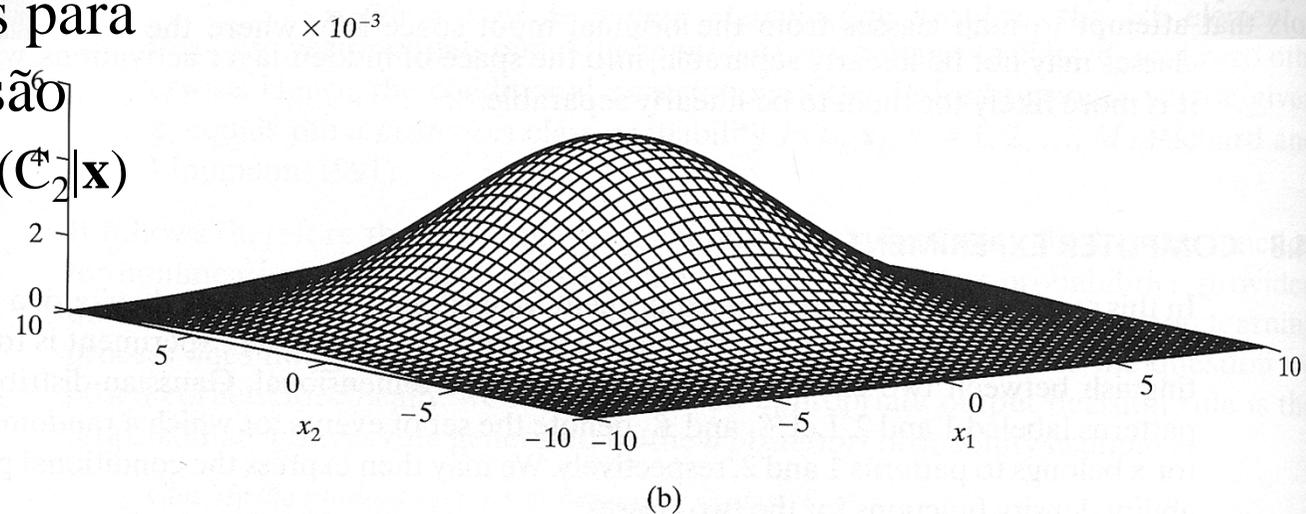
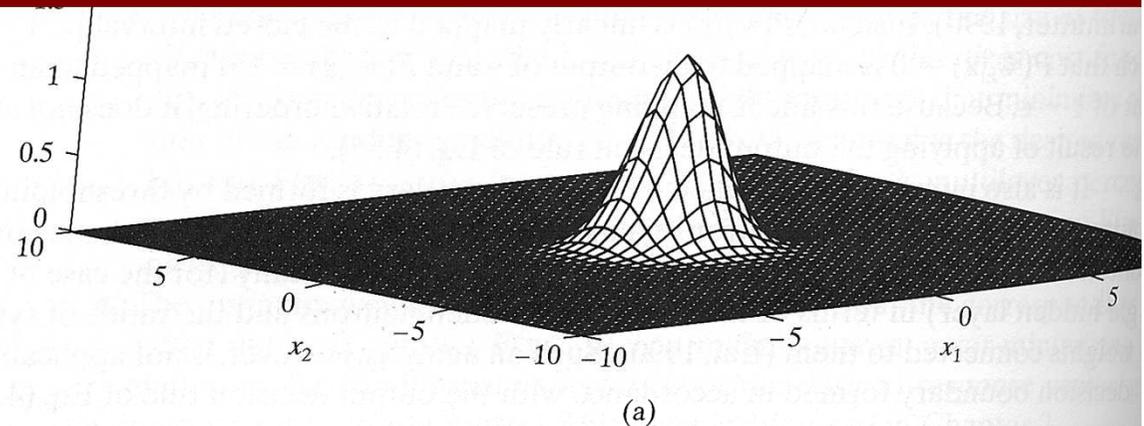


FIGURE 4.13 (a) Probability density function $f_x(\mathbf{x}|\mathcal{C}_1)$; (b) Probability density function $f_x(\mathbf{x}|\mathcal{C}_2)$.

Máquinas de Comitê Estáticas: Média do Conjunto - Exemplo

- Forma da fronteira de decisão: Circular com centro em $[-2/3, 0]$

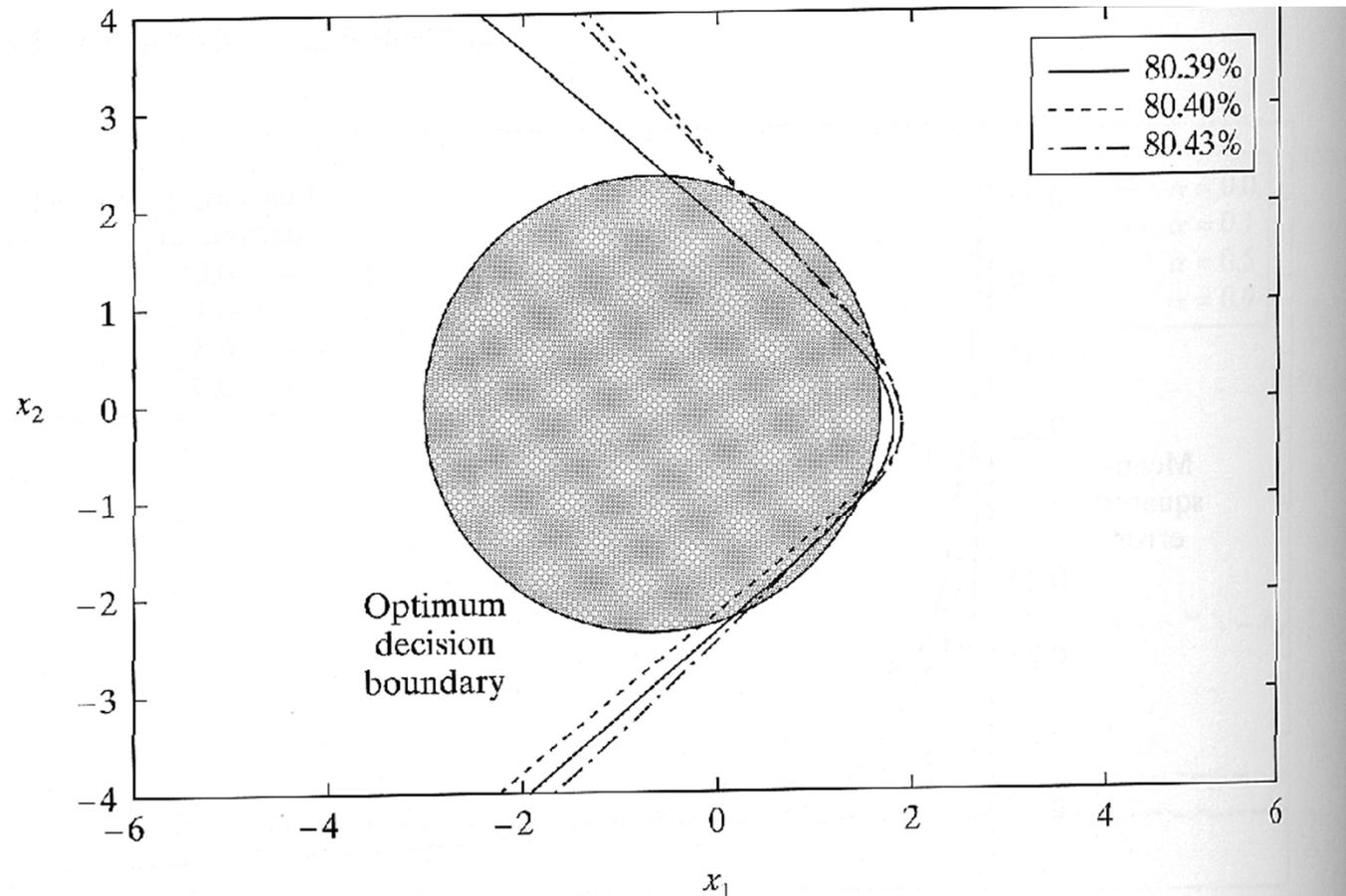


FIGURE 4.17A Plot of three "best" decision boundaries for the classification accuracies: 80.39, 80.40, and 80.43%.

Máquinas de Comitê Estáticas: Média do Conjunto - Exemplo

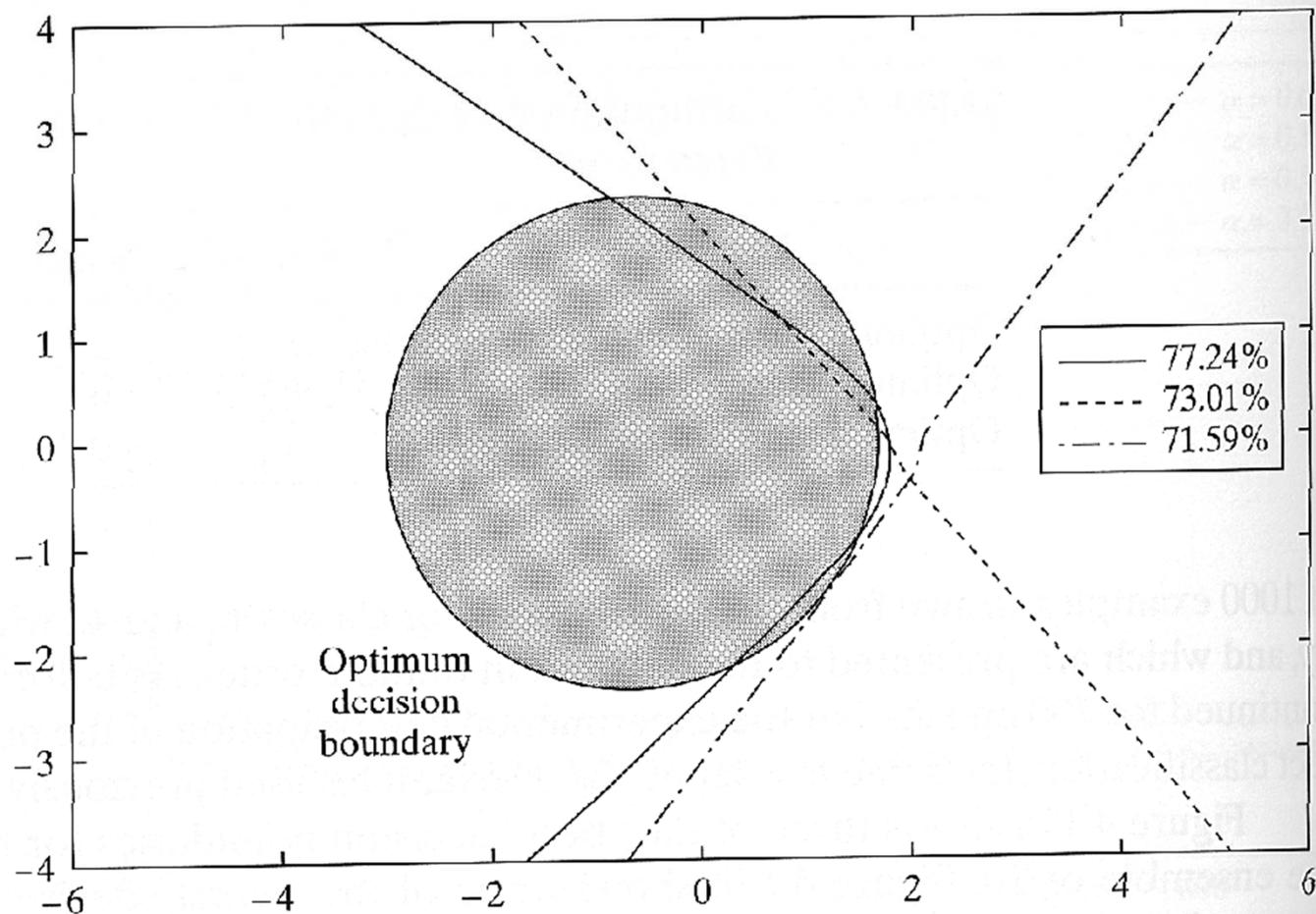


FIGURE 4.17B Plot of three "poorest" decision boundaries for the following classification accuracies: 77.24, 73.01, and 71.59%.

Máquinas de Comitê Estáticas: Média do Conjunto - Exemplo

- Combinando as saídas das 10 redes (especialistas), a média do conjunto tem erro esperado menor que o erro esperado da média dos erros individuais de cada especialista com respeito ao conjunto de dados considerado:
 - $79.4\% < 80.3\%$

TABLE 7.1 Classification Performances of Individual Experts L in a Committee Machine

Expert	Correct classification percentage	
Net 1	80.65	} média do conjunto: 79.4
Net 2	76.91	
Net 3	80.06	
Net 4	80.47	
Net 5	80.44	
Net 6	76.89	
Net 7	80.55	
Net 8	80.47	
Net 9	76.91	
Net 10	80.38	
Média	80.3	

Máquinas de Comitê Estáticas: Média do Conjunto - Ensacando

- Os especialistas resultantes são parecidos pois são treinados com conjunto de dados levemente diferentes entre si em contraponto à média de conjunto onde os dados de treinamento são os mesmos.
- *Bagging - Bootstrap Aggregating*) - Breiman, 1996:
 - Selecione um conjunto de treinamento D , com tamanho N
 - Para cada especialista (rede, árvore, K -vizinhos, etc)
 - Construa um novo conjunto de treinamento escolhendo N amostras, com reposição, de D .
 - Treine o especialista com o novo conjunto de dados.
 - Fim-do-Para
 - Saída é média ou votação de todas as máquinas treinadas.



Máquinas de Comitê Estáticas: Ensacando - Exemplo

- Resultados (percentual de erros) para dados de UCI (10-
validação cruzada)

	<i>Single net</i>	<i>Simple ensemble</i>	<i>Bagging</i>
<i>breast cancer</i>	3.4	3.5	3.4
<i>glass</i>	38.6	35.2	33.1
<i>diabetes</i>	23.9	23	22.8

- Desempenho com variações relevantes para pequenas
mudanças no conjunto de treinamento.

Máquinas de Comitê Estáticas: Impulso (*Boosting*)

- No impulso, há subconjuntos diferentes de dados, com distribuições diferentes, que são treinados em redes “fracas” (desempenho pouco superior a 50%), de modo que elas se especializem em partes distintas do conjunto de dados.
 - Tentativa de gerar aprendizes de base fracos que sejam complementares por treinamento de um aprendiz considerando os erros de aprendizes anteriores.
 - O objetivo é conversão em modelo de aprendizagem forte (acuracidade arbitrária com probabilidade alta).
- Aplicável a qualquer máquina de aprendizagem, podendo melhorar seu desempenho.

Máquinas de Comitê Estáticas: Impulso (*Boosting*)

- Três são os métodos para implementar a Máquina de Comitê por Impulso:
 - Filtragem: Utiliza grande fonte de exemplos que podem ser usados ou descartados durante o treinamento.
 - Subamostragem: Emprega conjunto de treinamento de tamanho fixo que são re-amostrados durante o treinamento de acordo com uma distribuição de probabilidade.
 - Re-ponderação: Emprega conjunto de treinamento fixo no qual os exemplos são ponderados por um algoritmo de aprendizado fraco.

Máquinas de Comitê Estáticas: Impulso (*Boosting*)

- Pseudo-código geral do método:

Selecione um conjunto de treinamento D com N amostras
Faça M vezes

Treine uma rede para o conjunto D

Encontre os exemplos em D que geram erro na rede

Dê ênfase nestes padrões e retire dos demais, criando
um novo conjunto de treinamento D_2

Atualize $D=D_2$

Fim-do-faça

Saída é média ou votação dos especialistas treinados

Máquinas de Comitê Estáticas: Impulso (*Boosting*) - Filtragem

- Envolve filtrar os exemplos de treinamento por diferentes versões de um algoritmo de aprendizagem fraca:
 - Assume grande disponibilidade de exemplos;
 - Exemplos são descartados ou mantidos durante o treinamento;
 - Requer pouca memória.
- Algoritmo proposto por Schapire (1990):
 - Fundamentado em um modelo de aprendizagem de distribuição independente ou *provavelmente aproximadamente correto* (PAC);
 - Na PAC, uma máquina de aprendizagem visa encontrar uma hipótese ou regra de previsão com uma taxa de erro máxima igual a ϵ (*modelo de aprendizagem forte*).

Máquinas de Comitê Estáticas: Impulso - Filtragem (Shapire, 1990)

- Treine Especialista 1 (E_1) com um conjunto de exemplos N_1 de tamanho N ;
- Filtre os dados através de E_1 para obter o segundo conjunto de dados, N_2 :
 - Jogue uma moeda, se der cara
 - Passe os dados de N_1 por E_1 até haver um erro, adicione este ponto a N_2
 - Se der coroa
 - Descarte os incorretos até achar o primeiro correto, adicione-o a N_2
 - Repita até N_2 ter o o tamanho de N

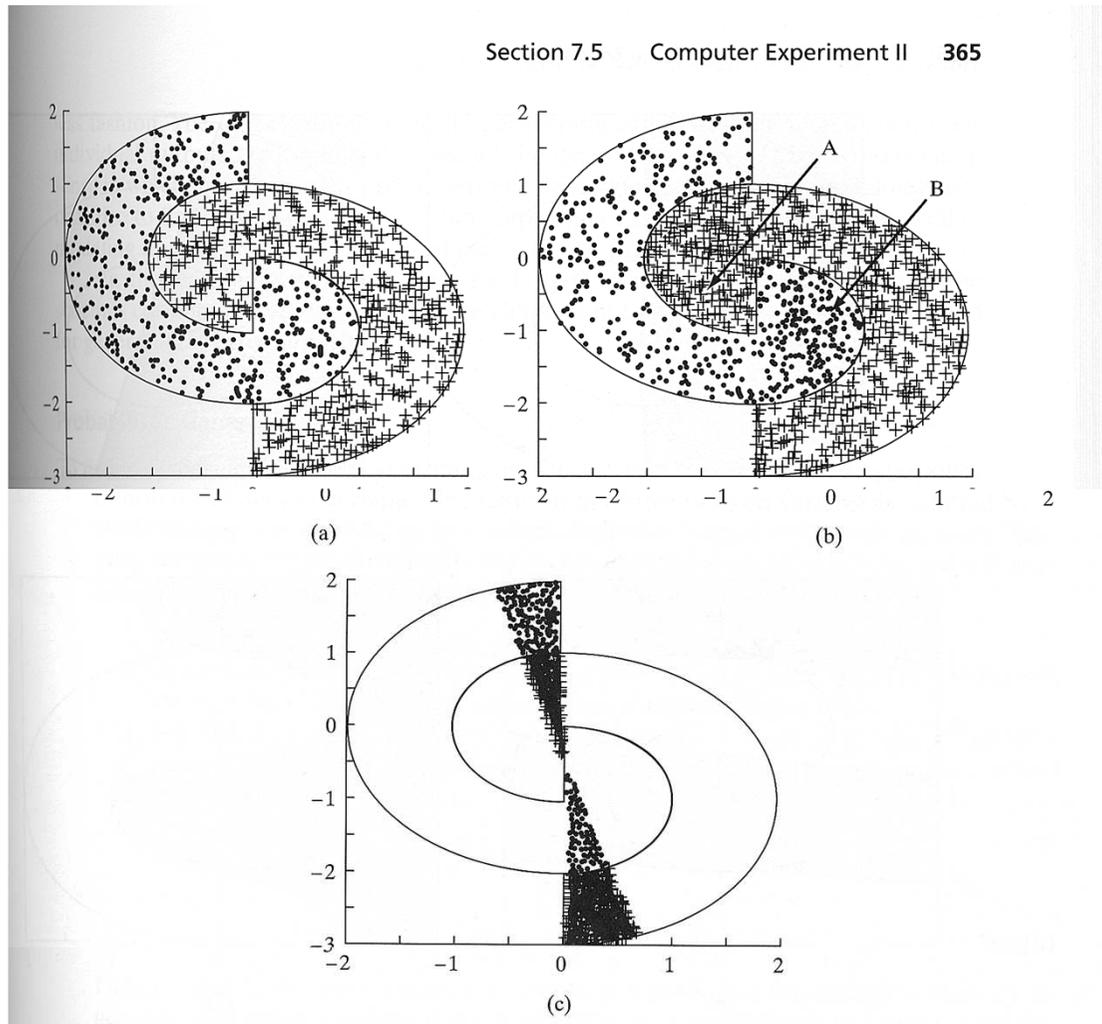
OBS: Se o Especialista 1 for testado com N_2 então a distribuição das entradas é tal que deve-se chegar a 50% de respostas corretas.

Máquinas de Comitê Estáticas: Impulso - Filtragem (Shapire, 1990)

- Treine Especialista 2 (E_2) com o conjunto de exemplos N_2 e empregue os dois especialistas para gerar um novo conjunto de treinamento N_3 ;
- Filtre os dados através de E_1 e E_2 :
 - Se as saídas concordarem, descarte o padrão;
 - Se discordarem, adicione o padrão a N_3 ;
- Treine Especialista 3 (E_3) com o conjunto de exemplos N_3 ;
- A saída do comitê é formada pela adição das respostas geradas por cada especialista

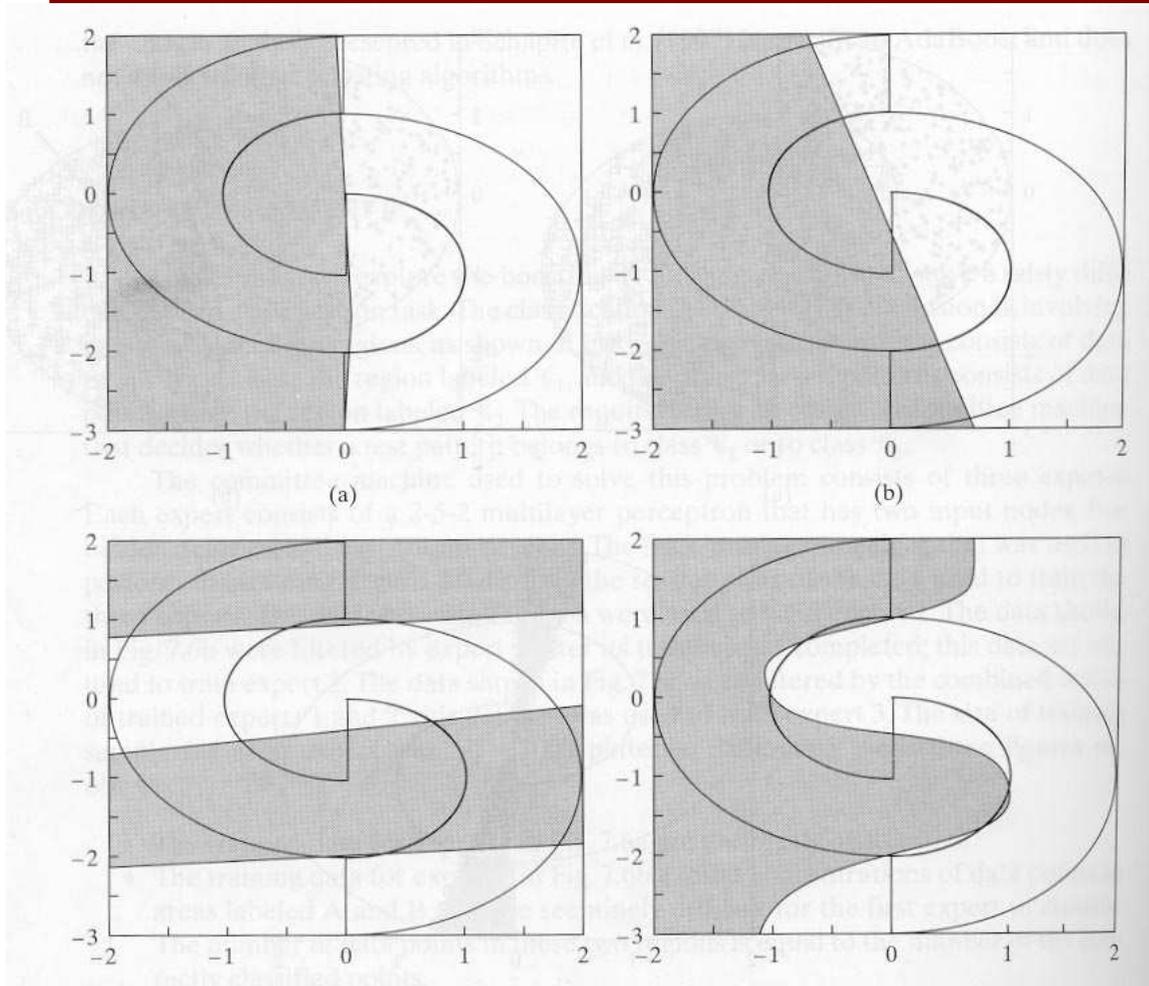
OBS: Os conjuntos N_2 e N_3 contém padrões de difícil aprendizagem pois o desempenho dos especialistas é $> 50\%$.

Máquinas de Comitê Estáticas: Impulso - Filtragem - Exemplo



- Classificação: Fronteiras são as linhas sólidas: pontos e cruzeiros em classes diferentes.
- Na figura, a distribuição dos conjuntos de dados: (a) Dados de treinamento (N_1), para Especialista 1 (E_1), são uniformemente distribuídos; (b) N_2 para E_2 concentra mais pontos em A e B, difíceis para E_1 classificar; e (c) N_3 para E_3 concentra dados de classificação difíceis para E_1 e E_2 .

Máquinas de Comitê Estáticas: Impulso - Filtragem - Exemplo



- Probabilidade de classificação correta:
 - (a) E1: 75,15%
 - (b) E2: 71,44%
 - (c) E3: 68,90%
 - (d) Comitê: 91,79%

Máquinas de Comitê Estáticas: Impulso por Sub-amostragem

- *AdaBoost: Adaptive Boosting.*
- Os exemplos são re-amostrados durante o treinamento, de acordo com uma dada distribuição de probabilidade.
- O erro é calculado em relação à amostra de treinamento fixa.
- O modelo de impulso por filtragem demanda grande quantidade de dados, se não estão disponíveis, a solução foi dada pelo impulso por sub-amostragem:
 - Modelo popular: Adaboost (Shapire, 1996)
 - Esta é uma variação do impulso por filtragem.

Máquinas de Comitê Estáticas: Impulso por Sub-amostragem - AdaBoost

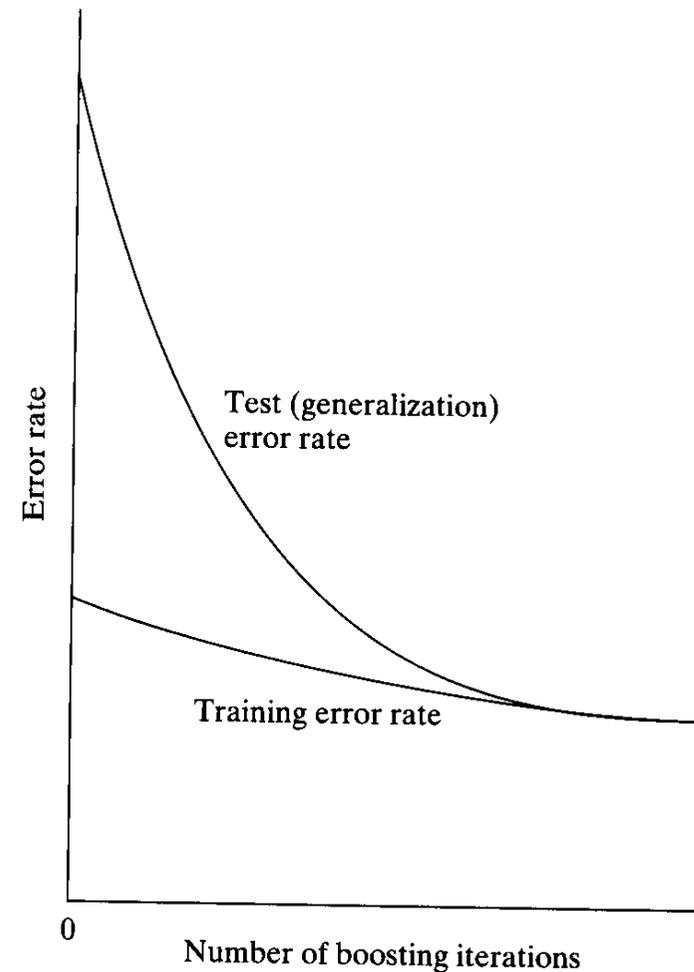
- Modelo não demanda grande quantidade de padrões de treinamento pois permite a reutilização destes dados:
 - Emprega conjuntos de dados de tamanho fixos.
- A estrutura de amostragem do AdaBoost é a estrutura natural da aprendizagem por lote (*batch*).
- Emprega modelo de aprendizagem fraca:
 - Ajusta adaptativamente o erro de hipóteses fracas vindas de modelos de aprendizagem fraca.
 - Seu desempenho depende daquele do modelo de aprendizagem fraca para as distribuições empregadas na aprendizagem.
- Quando o número de possíveis classes é superior a 2, o problema de impulso se torna mais complicado.

Máquinas de Comitê Estáticas: Impulso por Sub-amostragem - AdaBoost

- Algoritmo é caracterizado por
 - Gerar em cada passo (de forma determinística, mas adaptativa) uma distribuição sobre as amostras;
 - Dar maior peso (maior probabilidade de estar na amostra perturbada) às observações classificadas incorretamente no passo anterior.
- Forma de operar do algoritmo:
 - A cada tempo, k , o algoritmo provê um amostra \mathbf{x} de treinamento para a rede retirada do conjunto de amostras X empregando uma distribuição de probabilidade D_k usada para treinar a hipótese h_k .
 - O processo dura T instantes de tempo, depois dos quais o algoritmo usa média ponderada para combinar as saídas de T especialistas.
 - A distribuição D_{k+1} é calculada a partir de D_k , decrescendo a probabilidade de uma entrada ser escolhido se h_k foi corretamente classificada, focando nos padrões de mais difícil classificação.

Máquinas de Comitê Estáticas: Impulso por Sub-amostragem - AdaBoost

- A curva de erro do AdaBoost tem variação peculiar:
 - Sua forma depende da definição da confiança.



Máquinas de Comitê Estáticas: AdaBoost - Exemplo

- Resultados para dados de UCI (10-validação cruzada)

	<i>Single net</i>	<i>Simple ensemble</i>	<i>Bagging</i>	<i>AdaBoost</i>
<i>breast cancer</i>	3.4	3.5	3.4	4
<i>glass</i>	38.6	35.2	33.1	31.1
<i>diabetes</i>	23.9	23	22.8	23.3

- Fonte: Opitz & Maclin, 1999

Máquinas de Comitê Estáticas: Impulso por Ponderação

- Também trabalha com uma amostra de treinamento fixa.
- Porém assume que o algoritmo de aprendizagem fraca pode receber exemplos “ponderados”.
- O erro é calculado em relação aos exemplos ponderados.

Máquinas de Comitê Dinâmicas:

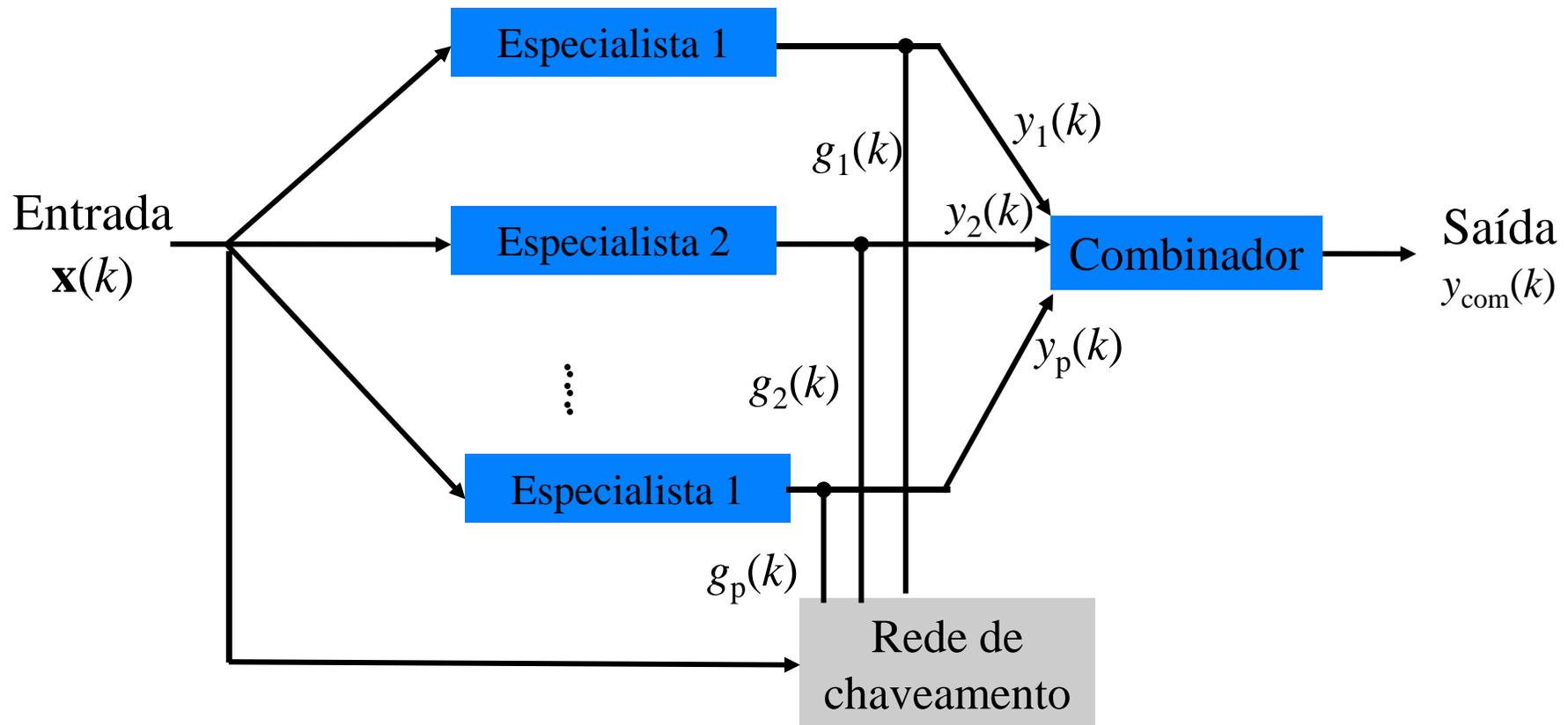
- Nas máquinas de comitê dinâmicas, o sinal de entrada está diretamente envolvido na combinação das saídas.
- Uma ou mais redes de chaveamento decide o peso de cada especialista.
- Considere uma rede na qual o processo de aprendizagem acontece fundindo de um modo suave as formas auto-organizada e supervisionada.
 - Os especialistas se auto-organizam para encontrar uma boa partição do espaço de entrada, de modo que cada especialista modele bem seu próprio subespaço.

Máquinas de Comitê Dinâmicas:

- Tipos de Estruturas:

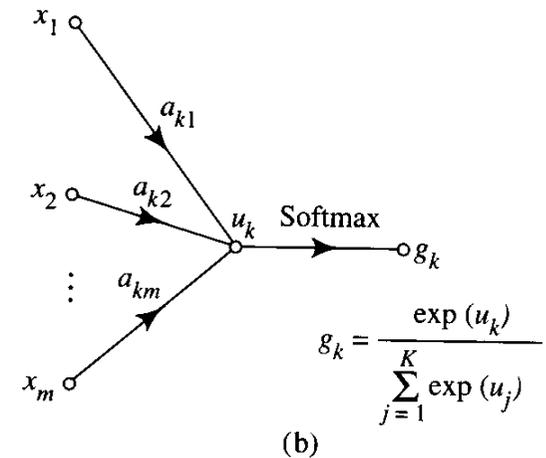
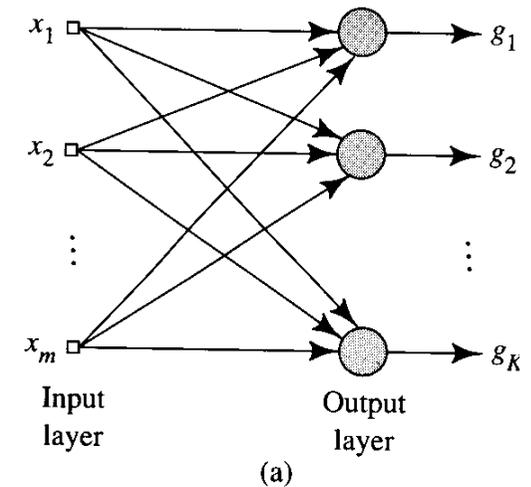
- Mistura de Especialistas (*mixtures of experts*):
- Mistura Hierárquica de Especialistas (*hierarchical mixtures of experts*):

Máquinas de Comitê Dinâmicas



Máquinas de Comitê Dinâmicas: Mistura de Especialistas

- A rede de chaveamento consiste de uma camada simples com K nodos, cada um para um especialista específico:
 - Seus nodos são não-lineares:
 - $y_k = g_k = \exp(u_k) / \sum_{j=1}^K \exp(u_j)$,
 $k = 1, 2, \dots, K$ and $u_k = \mathbf{a}_k^T \mathbf{x}$
 - g_k : probabilidade do especialista k estar correto.
- A rede de chaveamento é um classificador que mapeia a entrada \mathbf{x} para probabilidades multimodais:
 - Diferentes especialistas podem atingir a resposta desejada.



Máquinas de Comitê Dinâmicas: Mistura Hierárquica de Especialistas

• HME é uma extensão de ME. O primeiro se diferencia por ter seu espaço de entrada dividido em um conjunto aninhados. A informação é combinada e re-distribuída entre os especialistas sob controle de algumas redes de chaveamento dispostas hierarquicamente.

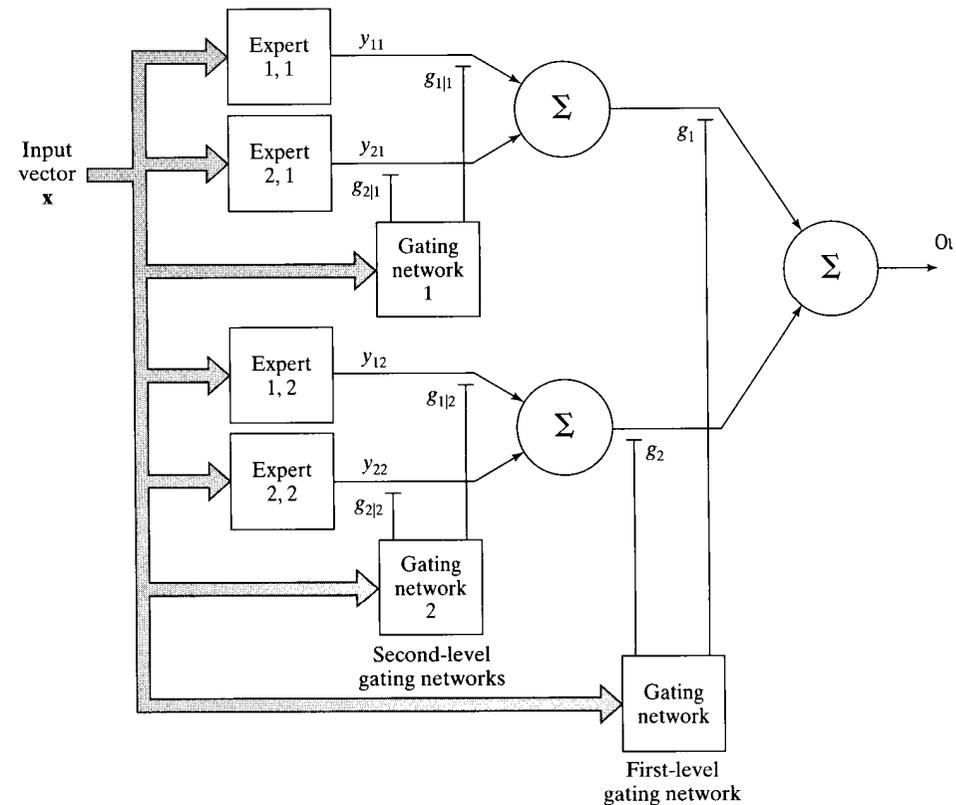


FIGURE 7.11 Hierarchical mixture of experts (HME) illustrated for two levels of hierar

Referências

- Haykin, S. (2001). *Redes Neurais, Princípios e Prática*. Segunda Edição.
- David Opitz, D. & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11: 169-198.
- Sharkey, A. J. C. (1996). On combining artificial neural nets. *Connection Science*, 8 (3-4): 299 - 314.
- Tresp, V. (2001). Committee machines. Chapter in *Handbook for Neural Network Signal Processing*, Yu Hen Hu and Jenq-Neng Hwang (eds.), CRC Press.