

Universidade de São Paulo

Escola de Artes, Ciências e Humanidades

Disciplina: SIN 5016 – Aprendizado de Máquina

Docente: Prof. Dr. Clodoaldo A Moraes Lima

Discentes: _____ No. USP: _____

1ª Questão) (1.0 ponto) Com relação aos modelos lineares, pede-se

- a) Cite uma vantagem dos modelos lineares em relação aos modelos não lineares;
- b) Explique detalhadamente os modelos lineares regularizados;
- c) Explique como podemos obter modelos não lineares a partir de modelos lineares;
- d) Qual a diferença entre modelos lineares e regressão logística.
- e) O que significa dimensão VC? Qual a dimensão de um hiperplano?

2ª Questão) (1.0 Ponto) Com relação a comitê de máquinas, pede-se

- a) Por que se diz que os componentes de um ensemble devem divergir em relação ao erro? Apresente alguma técnica para gerar diversidade de comportamento entre modelos de aprendizado voltados para a solução de um mesmo problema. Explique por que esta divergência no erro conduz ao melhor desempenho.
- b) Em que consiste as operações de geração, seleção e combinação em ensembles?
- c) Por que, mesmo havendo um componente que apresente um desempenho individual superior ao desempenho do ensemble, ainda assim é vantajoso adotar um ensemble?
- d) Quais são as principais distinções entre um ensemble e uma mistura de especialistas?
- e) Apresente um pseudo-código, com as respectivas equações, para treinamento de uma mistura de especialista gated (especialistas e gating como mlp) usando treinamento simultâneo e algoritmo EM.

3ª Questão) (1.0 Ponto) Com relação a deep learning, pede-se

- a) O que são hiperparâmetros em deep learning e por que é tão relevante definir valores adequados para eles?
- b) Por que se diz que as camadas convolucionais, seguidas ou não por operações de sub-amostragem, são extratores de características? Apresente os benefícios de filtros de sub-amostragem.
- c) O que é padding e o que é strider em uma convolução?
- d) Dada uma imagem $N \times M$, se aplicarmos a operação de convolução usando um filtro $p \times p$ com strider q , qual será o tamanho da imagem gerada.
- e) Dado uma CNN pretreinada para reconhecimento facial. Como podemos utilizá-la para extração de características faciais?
- f) Como calcular os pesos em um treinamento ao empregar dropout? Por que dropout pode ser considerado uma forma de regularização?

4ª Questão) (1.0 ponto) Com relação a autoencoder, pede-se

- a) O que é um autoencoder básico? Qual a finalidade do autoencoder?

- b) Como autoencoders podem ser utilizados na etapa de pré-treinamento de uma rede neural profunda? Qual a vantagem desta abordagem? **Dica: Leia o artigo abaixo**
- c) A figura 1, ilustra um tipo específico de autoencoder.
- b1) Explique detalhadamente este autoencoder
- b2) Mostre qual a vantagem da reparametrização na aplicação do backpropagation;
- b.3) Explique como esta abordagem pode ser usada para modelos generativos. Quais a diferença entre GAN e GMN?

https://www.ni.tu-berlin.de/fileadmin/fg215/teaching/nnproject/cnn_pre_trainin_paper.pdf

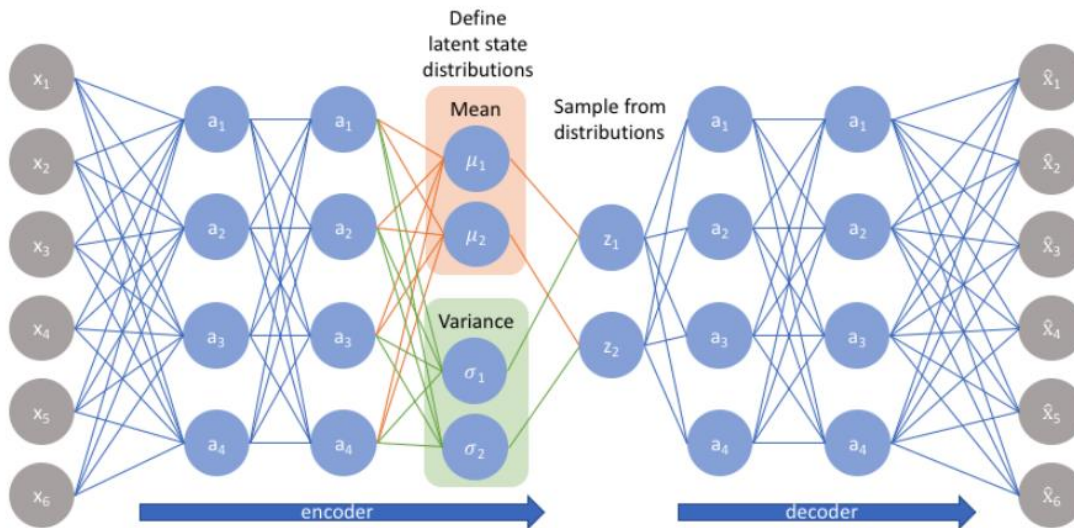


Figura 2

5ª Questão (1,0 Ponto) Considere o problema de aprender o conceito sobre pacientes doentes (+) ou saudáveis (-). Foram coletados os seguintes exemplos

Conjunto de treinamento			
Idade	Exame A	Exame B	Classe
21	H	C	+
25	R	B	+
28	R	C	+
24	H	B	+
19	J	B	+
35	R	C	-
30	J	C	-
38	J	B	-
40	H	C	+
25	J	C	-
29	R	C	-
20	J	B	-
19	R	C	-

Conjunto de Teste			
20	R	B	+
22	J	B	-
30	J	C	-
31	R	C	+
29	H	C	+

- a) Utilizando o algoritmo de indução de árvores de decisão, construa a árvore correspondente (sem poda e sem número mínimo de exemplos em cada folha), utilizando o critério de ganho de informação para selecionar atributos para este conjunto de exemplos. Anote, para cada nível da árvore de decisão, o valor do ganho de informação calculado para cada atributo, bem como aquele escolhido para particionar os exemplos. Se houver

empate entre valores do ganho de informação, escolha o primeiro (na ordem da tabela acima). Se for necessário realizar a discretização dos atributos, descreva detalhadamente o processo adotado.

- b) Análise de desempenho: use a árvore de decisão produzida em (a) para classificar os exemplos do conjunto de teste. Informe a precisão e matriz de confusão da árvore para esses exemplos. Discuta resumidamente os resultados.
- c) Aplique duas critérios de poda visto em sala de aula a árvore gerada no item b). Em seguida apresente a matriz de confusão para o conjunto de teste.
- d) Aplique duas critérios de poda visto em sala de aula a árvore gerada no item b). Em seguida apresente a matriz de confusão para o conjunto de teste.

6ª Questão) (1.0 ponto) Considerando Máquinas de Vetores Suporte, responda os itens abaixo.

- a) Descreva detalhadamente a SVM e suas variações para o problema de classificação (pelo menos duas abordagens). Apresente a formulação do primal e do dual. A solução obtida por um classificador linear regularizado é a mesma obtida por SVM linear, justifique sua resposta.
- b) Descreva detalhadamente como a SVM pode ser empregado para problemas com múltiplas classes. Apresente três estratégias e explique detalhadamente cada uma.
- c) Explique detalhadamente os três métodos alternativos para treinamento da SVM (Chunking, Algoritmo de Osuna, SMO).
- d) Explique o que significa o Truque do Kernel (Kernel Trick). Como podemos transformar uma abordagem linear em não linear.
- e) Explique detalhadamente como podemos aplicar dimensão VC para selecionar o tipo de kernel.

7ª Questão) (1.0 ponto) Com relação a redes neurais artificiais, responda os itens abaixo

- a) Apresente um pseudo-código para o algoritmo Back-Propagation, considerando a correção do erro em lote e sequencial. O aluno deve apresentar as equações de cálculo da saída da rede, do gradiente e atualização dos pesos, considerando função de custo entropia cruzada, função de ativação na camada escondida relu e na saída softmax.
- b) Descreva quais hiperparâmetros o usuário deve definir para executar uma Rede Neural MLP. Apresente uma estratégia para definição de cada um desses parâmetros.
- c) Com relação aos algoritmos de otimização irrestrita, apresente duas vantagens e duas desvantagens dos algoritmos de segunda e primeira ordem para atualização dos pesos. Por que os métodos de segunda conduzem a convergência mais rápida?
- d) Alguns autores sugerem inserir um termo de penalidade na função objetiva para controlar a suavidade do mapeamento produzido pela RNA. Descreva detalhadamente esta abordagem e apresente a função objetiva a ser minimizada. Qual o significado do parâmetro de regularização? Dado que termo de momento consiste

em adicionar uma expressão na função objetiva, este pode ser considerado uma forma regularização? Justifique detalhadamente sua resposta.

- e) Faça um estudo comparativo entre Redes Neurais Multilayer Perceptron e Redes Neurais de Funções de Base Radial. Apresente **pelo menos duas** semelhanças e duas diferenças entre RBF e MLP. Além disso, discuta sobre os parâmetros a serem determinados em cada arquitetura.
- f) Explique o que é capacidade de generalização em treinamento supervisionado de redes neurais artificiais e como a disponibilidade de um conjunto de treinamento e de um conjunto de validação pode ser empregada visando maximizar esta capacidade.
- g) Explique o que é maldição da dimensionalidade. Como a maldição da dimensionalidade afeta o desempenho de um modelo neural. Esta maldição afeta também as máquinas de vetores suporte? Explique detalhadamente.

8ª Questão) (1.0 Ponto) Um dos papéis do estado-da-arte em qualquer área em expansão tecnológica é servir de referência para propostas ainda melhores, que então vão se tornar o estado-da-arte, numa evolução continuada. Este curso não tinha objetivo de atingir o estado-da-arte em deep learning, em qualquer frente que se queira considerar, seja pela imprevisibilidade da evolução tecnológica de ponta, seja pelo fato do curso estar voltado para aspectos mais conceituais e introdutórios. Cabe, no entanto, uma atividade conceitual voltada para algumas técnicas que conduzem a modelos de aprendizado estado-da-arte. Para tanto, o aluno deve descrever os aspectos principais das metodologias que conduzem à GoogLeNet (módulo inception), Resnet, EfficientNet e à FixEfficientNet, a partir dos seguintes textos da literatura:

<https://arxiv.org/pdf/1905.11946.pdf>

<https://arxiv.org/pdf/2003.08237.pdf>


<https://arxiv.org/pdf/1512.03385.pdf>

https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf

9ª Questão) (2,0 Ponto) Considere a imagem de um gato ilustrada na figura 1a) em RGB, com os respectivos canais. Seja uma Rede Neural Convolutacional composta pelas seguintes camadas

- 1ª camada) camada de convolução com 3 filtros com dimensão 2x2 strider 1, função de ativação sigmoid,
- 2ª camada) uma camada pooling 2x2 (pooling média) strider 2,
- 3ª camada) uma camada de convolução com 02 filtro 2x2 strider 2, função de ativação relu
- 4ª camada) uma camada pooling 2x2 (pooling máximo) strider 1
- 5ª camada) totalmente conectada com função ativação softmax e 10 saídas

- a) Faça um desenho ilustrando a rede convolutacional.
- b) Inicialize os valores dos filtros, camada totalmente conectada e calcule a saída da rede convolutacional
- c) Sabendo que o gato foi rotulado como classe desejada 1, retropropague o erro, calcule o valor do gradiente em relação aos coeficientes do filtro e o atualize usando taxa de aprendizado igual 0.1;

	<pre> 249 252 255 241 242 255 255 255 255 255 218 190 230 180 223 255 255 255 255 255 229 110 72 91 211 255 255 255 255 255 238 121 107 115 193 255 255 255 255 255 251 133 146 170 175 255 255 255 255 255 251 130 134 117 111 187 246 255 255 255 255 162 117 69 88 103 150 250 255 255 255 233 154 83 82 81 69 144 254 255 255 250 163 105 91 83 93 74 192 255 255 255 160 93 72 113 113 77 97 240 255 251 125 38 38 65 50 53 69 135 255 233 171 151 140 130 125 129 143 160 </pre>
a) Imagem Colorida (Original)	b) Canal R
<pre> 249 252 255 240 242 255 255 255 255 255 216 187 229 174 221 255 255 255 255 255 226 101 65 83 207 255 255 255 255 255 238 114 99 107 191 255 255 255 255 255 251 131 144 167 172 255 255 255 255 255 251 126 132 112 103 181 246 255 255 255 254 157 113 64 80 94 144 250 255 255 255 233 152 81 74 73 60 140 255 255 255 250 161 101 84 76 83 62 188 255 255 255 158 87 66 105 103 65 90 239 255 251 121 30 30 58 43 44 57 128 255 233 168 146 134 125 121 126 139 155 </pre>	<pre> 249 252 255 240 242 255 255 255 255 255 217 186 229 172 219 255 255 255 255 255 226 97 63 78 206 255 255 255 255 255 238 113 93 99 190 255 255 255 255 255 251 132 145 166 171 255 255 255 255 255 251 126 132 112 100 179 246 255 255 255 254 155 112 64 78 91 141 249 255 255 255 234 153 80 71 68 56 138 255 255 255 250 162 100 81 73 79 53 184 255 255 255 159 85 62 101 96 57 85 239 255 251 122 26 26 56 40 39 49 124 255 233 168 145 133 124 120 124 137 155 </pre>
c) Canal G	d) Canal B
Figura 1	