

CSCI 416/516 Final Study Guide

Name:

Student ID:

1 Support Vector Machine

- **Problem 1: Theory of SVMs.**

Explain the principle of maximizing the margin in SVMs. How does this contribute to the model's generalization ability?

- **Problem 2: Kernel Tricks.**

What is the kernel trick in SVMs? Provide examples of different types of kernels used in SVMs.

- **Problem 3: Support Vectors.**

Define support vectors and explain their significance in the context of SVMs.

- **Problem 4: Parameter Tuning.**

Discuss the role of parameters like C in SVMs. How do they affect the model's performance?

- **Problem 5: Hyperplane Decision Boundary.**

What is a hyperplane in SVMs, and how does it help in classification tasks?

- **Problem 6: Dual Problem and Kernel Trick.**

Explain how the dual problem in SVMs facilitates the use of the kernel trick. Why is this significant?

- **Problem 7: Lagrangian Multipliers.**

Suppose a sample in the training dataset has a Lagrangian multiplier being 0.1. What does this say about this sample?

- **Problem 8: Primal and Dual.**

How is the Primal in SVMs defined? And how is it related to the Dual?

- **Problem 9: Linear Separability.**

Describe the techniques or strategies you can use to address the issue of linear inseparability when working with SVMs.

- **Problem 10: Kernel Tricks.**

What are the advantages and disadvantages of the kernel tricks?

2 Decision Tree

	Cloudy	Not Cloudy
Raining	24/100	1/100
Not Raining	25/100	50/100

- **Problem 11: Joint Entropy.**

Suppose $X = \{\text{Raining, Not raining}\}$, $Y = \{\text{Cloudy, Not cloudy}\}$. What is the joint entropy, $H(X, Y)$?

- **Problem 12: Specific Conditional Entropy.**

Following the setup from the previous question, what is the entropy of cloudiness Y , given that it is raining ($X = \text{raining}$)?

- **Problem 13: General Conditional Entropy.**

Following the setup from the previous question, what is the entropy of cloudiness Y , given the variable X ?

- **Problem 14: Information Gain.**

How is the Information Gain $IG(Y|X)$, given the entropy of Y , $H(Y)$, and the conditional entropy $H(Y|X)$?

- **Problem 15: Information Gain.**

if X is completely uninformative about Y , what is the value of $IG(Y|X)$?

- **Problem 16: Information Gain.**

if X is completely informative about Y , what is the value of $IG(Y|X)$?

- **Problem 17: Overfitting.**

What is overfitting in the context of decision trees?

- **Problem 18: Tree Components.**

Explain the concepts of nodes, branches, leaves, and root node in a decision tree.

- **Problem 19: IG and Tree.**

How is the Information Gain used to build a decision tree?

- **Problem 20: Pros and Cons.**

What are the advantages and disadvantages of using the decision tree?

3 Ensemble Learning

- **Problem 21: Weak Learners.**

What are weak learners in the context of AdaBoost? Provide examples of common weak learners.

- **Problem 22: Misclassification.**

Discuss the concept of misclassification rate and its role in AdaBoost. How are weights adjusted for misclassified samples?

- **Problem 23: Decision Stumps.**

Discuss the relationship between AdaBoost and decision stumps. Why are decision stumps often chosen as weak learners in AdaBoost?

- **Problem 24: Sample weights.**

What does the weight $w_{t,i}$ of a given sample x_i mean in the context of AdaBoost?

- **Problem 25: Objective function.**

Explain what this objective function does, in the context of AdaBoost.

$$\mathcal{J}_{\text{reg}}(\boldsymbol{\theta}) = - \sum_{i=1}^n w_i [y_i \log h_{\boldsymbol{\theta}}(\mathbf{x}_i) + (1 - y_i) \log(1 - h_{\boldsymbol{\theta}}(\mathbf{x}_i))] + \lambda \|\boldsymbol{\theta}_{[1:d]}\|_2^2 \quad (1)$$

4 Multilayer Perceptrons

- **Problem 26: Backpropagation.**

Discuss the significance of the learning rate in the backpropagation algorithm. What can happen if it is set too high or too low?

- **Problem 27: Architecture.**

Describe the typical architecture of a Multilayer Perceptron. Include details about the input layer, hidden layers, and the output layer.

- **Problem 28: Learning Rate.**

Explain the significance of the learning rate in the training process of an MLP. What are the potential effects of setting it too high or too low?

- **Problem 29: Gradient Descent.**

What's the relationship between gradient descent and backpropagation?

- **Problem 30: Activation Functions.**

Explain the role of activation functions in MLPs.

5 Convolutional Neural Networks

- **Problem 31: Convolution.**

What is the value given a vector $\mathbf{a} = [2, -1, 1]$ convolved by a filter $\mathbf{b} = [1, 1, 2]$?

- **Problem 32: Activation Functions.**

Given the answer from the previous question, what is the outcome after you apply ReLU on it?

- **Problem 33: Kernels.**

What is the purpose of the filters/kernels used in a convolutional layer?

- **Problem 34: Overfitting.**

What methods can be used to prevent overfitting in CNNs?

- **Problem 35: Pooling.**

What is the purpose of pooling layers in a CNN? Compare and contrast Max Pooling and Average Pooling.

6 Attention & Transformers

- **Problem 36: Self-Attention.**

Explain the concept of self-attention in transformers.

- **Problem 37: Multi-Head Attention.**

What is multi-head attention in transformers?

- **Problem 38: Encoders and Decoders.**

Compare the roles of the encoder and decoder in a transformer model. How are they similar and different?

- **Problem 39: Attention.**

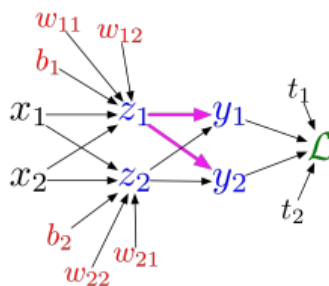
Describe how the concept of attention, as used in transformers, can be applied to domains other than language processing, such as image or video analysis.

- **Problem 40: Application.**

Suppose you want to use transformers for multiclass classification. How should you modify the existing transformer architecture to achieve this goal?

7 Miscallenous

Multiclass logistic regression



$$z_\ell = \sum_j w_{\ell j} x_j + b_\ell$$

$$y_k = \frac{e^{z_k}}{\sum_\ell e^{z_\ell}}$$

$$\mathcal{L} = - \sum_k t_k \log y_k$$

- **Problem 41: Backpropagation.**

What does the back pass look like, given the illustrated forward pass?