

به نام خدا



دانشکده مهندسی کامپیوتر

مبانی هوش محاسباتی

تمرین دوم

SVM – Kernel trick – Decision tree – boosting

دکتر مائده سادات طاهایی

زمستان 1403

طراحان تمرین: حسین الله وردی – یونس عبدالملکی – ارشیا حسین زاده



- در صورت وجود هرگونه ابهام به طراح پیام دهید.
- انجام تمرین ها تک نفره می باشد.
- زبان برنامه نویسی پایتون است.
- موارد ارسال شده به صورت آنلاین تحویل گرفته خواهند شد.
- کل فایل محتوای ارسالی را داخل فایل زیپ قرار داده و نام آن را شماره دانشجویی خود قرار دهید.
- تاریخ ریلیز تمرین: 25 فروردین ماه
- تاریخ تحویل تمرین: 8 اردیبهشت ماه
- آیدی طراحان در تلگرام: @arshia_hz80 younes_abdolmalaky@
@hossein1377a



1. سوالات تئوری

سوال 1 (آقای الله وردی):

الف) با توجه به مجموعه داده زیر که دارای جداسازی غیرخطی کلاس‌ها است، یک تابع کرنل بهینه پیشنهاد دهید و انتخاب خود را به صورت تحلیلی توجیه کنید. همچنین یک جدول مقایسه‌ای بین حداقل سه تابع کرنل از نظر پیچیدگی، تمایل به بیش‌برازش و قابلیت تفسیر ارائه دهید.

Random state: Student Number

جدول 1 - تحلیل مقایسه‌ای توابع کرنل

تابع کرنل	روابط	قابلیت تفسیر	تمایل به بیش‌برازش	پیچیدگی محاسباتی
خطی	$K(x_i, x_j) = x_i^T x_j$			
چندجمله‌ای	$K(x_i, x_j) = (\alpha x_i^T x_j + c)^d$			
RBF	$K(x, x') = e^{-\gamma x - x' ^2}$			

ب) تأثیر تغییر پارامتر تنظیم‌کننده C در SVM را بر توازن بایاس-واریانس تحلیل کنید. جدول زیر را برای نشان دادن این رابطه تکمیل کنید.

جدول 2 - اثر پارامتر تنظیم‌کننده (C) بر توازن بایاس-واریانس

مقدار C	واریانس	بایاس
0.01		
1		
100		

سوال 2) آقای الله وردی: در فضاها با ابعاد بالا، SVM ممکن است از مشکل ابعاد بالا

(curse of dimensionality) دچار خطا شود. یک استدلال تحلیلی ارائه دهید و در این باره تحقیق کنید.

سوال 3) آقای الله وردی: ماتریس‌های در هم ریختگی یک درخت تصمیم و یک مدل Boosting را که روی یک

مجموعه داده مشابه اعمال شده‌اند، اثبات و بررسی کنید. علت هرگونه اختلاف در عملکرد را شناسایی و توضیح دهید.

سوال 4) آقای الله وردی: یک جدول مقایسه تحلیلی بین مدل‌های درخت تصمیم، SVM و Boosting از نظر

مقیاس‌پذیری، قابلیت تفسیر و عملکرد روی داده‌های دارای نویز تهیه کنید.

جدول 3 - مقایسه پیشرفته مدل‌ها

مدل	عملکرد روی داده‌های دارای نویز	قابلیت تفسیر	مقیاس‌پذیری
SVM			
درخت تصمیم			
Boosting			

سوال 5) آقای عبدالمالکی: یک ماشین یادگیری AdaBoost روی داده‌ای به بزرگی $m=1000$ نمونه

آموزش دیده است. خطای آموزش $15/0$ بوده و نتیجه آموزش به شرح زیر است:

$$T=3, h_1(x)=2x_1+x_2-1, \alpha_1=0.3; h_2(x)=x_1-x_2+5, \alpha_2=0.35; h_3(x)=x_1-3x_2+1, \alpha_3=0.4$$

الف: برای $\delta=0.05$ و $\rho=0.6$ ، باند بالای خطای واقعی را حساب کنید.

ب: مقدار margin را برای $y=1$ و $x=(1, -1)$ حساب کنید.

سوال 6) آقای عبدالمالکی: الگوریتم بوستینگ را می توان به صورت مسأله بهینه سازی زیر نوشت:

مجموعه ای از ضرایب α به شکل زیر تعریف می شود:

$$\min_{\|\bar{\alpha}\|_1 \leq 1/\rho} \frac{1}{m} \sum_{i=1}^m e^{-y_i \bar{h}(x_i)} + \lambda \|\bar{\alpha}\|_1 \quad \bar{h}(x_i) = \sum_{t=1}^T \bar{\alpha}_t h_t(x_i)$$

الف: تابع $h(x)$ برای وقتی λ خیلی بزرگ است چگونه خواهد بود؟

ب: تابع $h(x)$ برای وقتی T خیلی بزرگ است چگونه خواهد بود؟

پ: تابع $h(x)$ برای وقتی λ خیلی کوچک است چگونه خواهد بود؟

ت: تابع $h(x)$ برای وقتی T خیلی کوچک است چگونه خواهد بود؟



سوال 7) آقای عبدالمالکی: . پس از آموزش SVM با استفاده از کرنل

$$K(x, x') = (1 + x' \cdot x)^2$$

نتایج زیر حاصل شده است. مقدار اتلاف (loss) را براساس تابع Hinge حساب کنید.

Support Vectors	Label	α
$x_1 = [0, 2]$	+1	0.5
$x_2 = [4, 0]$	+1	1.5
$x_3 = [1, 0]$	-1	0.5
$x_4 = [0, 0.5]$	-1	1.5

سوال 8) آقای عبدالمالکی: جدول زیر نتایج آموزش مدل SVM روی یک دیتاست را نشان می‌دهد که از کرنل

RBF استفاده شده است و $C = 2.5$ و $\gamma = 2$ است:

Support Vector (x)	Label (y)	α
$[1, 5, -\sqrt{21}]$	+1	22
$[1, 4, +2\sqrt{21}]$	-1	11
$[1, 5, +2\sqrt{21}]$	-1	2.5

میزان تخطی نمونه سوم را محاسبه کنید.

سوال 9) آقای عبدالمالکی: سه نمونه (example) مجموعه آموزش را تشکیل میدهد. داده ها یک ویژگی دارند. جدول زیر نمونه ها را معرفی کرده است.

داده	x	کلاس
1	-1	-1
2	0	+1
3	+1	-1

الف - با استفاده از تابع $K(x, x') = (x^T x' + 1)^2$ و مقدار $C = 3$ عملیات آموزش SVM را بنویسید.

ب - تابع $W^T x + b = 0$ را پیدا کرده و بر حسب x رسم کنید.

ج - عرض نوار را در فضای ویژگی حساب کنید.

سوال 10) آقای عبدالمالکی: به سوالات زیر در ارتباط با SVM پاسخ دهید:

الف) SVM ذاتاً برای مسائل دوتایی طراحی شده است. چگونه می توان SVM را برای مسائل چندکلاسه استفاده کرد؟ تفاوت روش های "One-vs-One" و "One-vs-All" در این زمینه چیست؟

ب) چرا تنها نقاط پشتیبان (Support Vectors) در تعیین مرز تصمیم گیری نقش دارند؟

پ) SVM چه تفاوت ها و شباهت هایی با الگوریتم های دیگری مانند Logistic Regression یا Neural Networks دارد؟ چرا SVM در برخی موارد ترجیح داده می شود؟

سوال 11) آقای حسین زاده): طبقه بندی با استفاده از SVM:

الف) در حالت تفکیک پذیر خطی، اگر یکی از نمونه‌های آموزشی حذف شود، آیا مرز تصمیم به سمت نقطه حذف شده حرکت می‌کند، از آن دور می‌شود، یا ثابت می‌ماند؟ پاسخ خود را توجیه کنید. حال اگر در نظر بگیریم که مرز تصمیم متعلق به Logistic Regression است، آیا مرز تصمیم تغییر می‌کند یا ثابت می‌ماند؟ پاسخ خود را توضیح دهید. (لازم نیست جهت تغییر را ذکر کنید).

ب) اگر اجازه دهیم مقدار کمی خطا در داده‌های آموزشی وجود داشته باشد، بهینه‌سازی اولیه SVM (soft margin) به صورت زیر است:

$$\begin{aligned} \min_{\mathbf{w}, \xi_i} \quad & 1/2 \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T(x_i)) \geq 1 - \xi_i, \forall i \in \{1, \dots, n\} \\ & \xi_i \geq 0, \forall i \in \{1, \dots, n\} \end{aligned}$$

فرض کنید مقادیر بهینه ξ_1, \dots, ξ_n محاسبه شده اند، از مقادیر ξ_i برای تعیین یک حد بالا (کران بالا) بر تعداد نمونه‌هایی که به درستی طبقه‌بندی نشده‌اند، استفاده کنید.

پ) در بهینه‌سازی اولیه SVM، ضریب C چه نقشی دارد؟ پاسخ خود را با در نظر گرفتن دو حالت حدی یعنی $C \rightarrow 0$ و $C \rightarrow \infty$ توضیح دهید.

ت) Hard SVM و Logistic Regression را در حالتی که دو کلاس به صورت خطی تفکیک پذیر باشند مقایسه کنید. تفاوت های مهم را بیان کنید.

ث) Soft SVM و Logistic Regression را در حالتی که دو کلاس به صورت خطی تفکیک پذیر نباشند مقایسه کنید. تفاوت های مهم را بیان کنید.

سوال 12 (آقای حسین زاده): جدول زیر 5 داده آموزشی در \mathbb{R}^1 را نشان می‌دهد. با استفاده از SVM و کرنل چندجمله‌ای درجه 2 آموزش انجام شده و نتیجه $h(x) = 0.66667x^2 - 5.33333x + b$ حاصل شده است.

الف) کدامیک از داده‌های بردارهای پشتیبان هستند؟

ب) مقدار b را حساب کنید.

پ) ضرایب وزن بردارهای پشتیبان (α) را حساب کنید.

توجه: چندجمله‌ای درجه 2 در \mathbb{R}^1 عبارت است از: $[x^2 \ \sqrt{2}x \ 1]$

x_1, y_1	x_2, y_2	x_3, y_3	x_4, y_4	x_5, y_5
1, +1	2, +1	4, -1	5, -1	6, +1

سوال 13 (آقای حسین زاده): جدول زیر شامل داده‌های آموزشی است که به پیش‌بینی احتمال حمله قلبی در بیماران کمک می‌کند:

Patient ID	Chest Pain	Male	Smokes	Exercises	Heart Attack
1	No	Yes	Yes	Yes	No
2	Yes	Yes	Yes	No	Yes
3	No	No	Yes	No	Yes
4	No	Yes	No	Yes	No
5	Yes	No	Yes	Yes	Yes
6	Yes	Yes	No	Yes	Yes

الف) با استفاده از آنترپی (Entropy) یک درخت تصمیم بهینه برای پیش‌بینی احتمال حمله قلبی ایجاد کنید. تمامی مراحل محاسبات را نشان دهید.

ب) بر اساس درختی که در بخش قبل ساخته‌اید، وضعیت یک بیمار جدید را که دچار درد قفسه سینه است، طبقه‌بندی کنید.

2. سوالات عملی

Exercise 1(Abdolmaleky): Implementing a Decision Tree from Scratch

Objective: Create a decision tree classifier to classify the Iris dataset.

Explanation:

1. Understanding the Dataset:

The Iris dataset contains four features (sepal length, sepal width, petal length, and petal width) and three target classes (species of iris: Setosa, Versicolor, and Virginica).

Familiarize yourself with the data and its structure.

2. Building the Decision Tree:

Node Structure :Define a structure to represent a node in the decision tree. Each node should have information about the feature used for the split, the threshold value, pointers to left and right child nodes, and a value representing the class label (for leaf nodes).

Splitting Criteria :Implement a method to calculate the impurity of a dataset (using Gini impurity or entropy). This method will help determine the best feature and threshold to split the data.

Recursive Splitting :Create a recursive function that splits the data based on the best feature and threshold until a stopping criterion is met (e.g., maximum depth, minimum samples per leaf, or if all instances at a node belong to the same class).

Prediction Method :Implement a method to traverse the tree for making predictions on new data points.

3. Training and Evaluation:

Split the dataset into training and testing sets.

Train your decision tree on the training set.

Evaluate the model's performance on the test set by calculating accuracy and other relevant metrics (precision, recall, F1-score).

Exercise 2: Implementing AdaBoost from Scratch(abdolmaleky)

Objective :Create an AdaBoost classifier that combines multiple weak classifiers to improve performance on the Iris dataset.

Explanation:

1. Understanding Boosting:

AdaBoost works by combining several weak classifiers (models that perform slightly better than random guessing) to create a strong classifier.

The algorithm focuses on misclassified instances from previous iterations, adjusting their weights so that future classifiers pay more attention to them.

2. Creating a Weak Classifier:

Define a simple weak classifier (e.g., a decision stump, which is a single-level decision tree). This classifier should be able to classify data points based on one feature and a threshold.

3. Implementing the AdaBoost Algorithm:

Initialization :Start by assigning equal weights to all training instances.

Training Iterations :For a specified number of iterations:

Train a weak classifier on the weighted training data.

Calculate the error rate of the classifier and its weight based on performance.

Update the weights of the training instances, increasing the weights for misclassified instances and decreasing the weights for correctly classified ones.

Final Classifier :Combine the predictions of all weak classifiers into a final strong prediction using the weighted votes from each classifier.

4. Training and Evaluation:

Split the dataset into training and testing sets.

Train your AdaBoost model on the training set.

Evaluate the model's performance on the test set by calculating accuracy and other relevant metrics.

Exercise 3: Implementing Support Vector Machine (SVM) from Scratch

Objective : Create a Support Vector Machine (SVM) classifier to classify the Iris dataset.

Explanation:

1. Understanding SVM:

Support Vector Machines aim to find the optimal hyperplane that separates data points of different classes in a high-dimensional space.

The points closest to the hyperplane are called support vectors, and they are critical in defining the decision boundary.

2. Data Preparation:

Load the Iris dataset and understand the features and target classes.

For simplicity, consider converting the problem to binary classification (e.g., classifying Setosa vs. Non-setosa).

3. Implementing the SVM Algorithm:

Hinge Loss Function :Implement a method to compute the hinge loss, which measures how well the current hyperplane separates the data points.

Gradient Descent :Use gradient descent to optimize the weights and bias of the hyperplane. Update the weights based on the hinge loss and learning rate.

Kernel Trick (Optional) :If desired, implement kernel functions (like polynomial or Gaussian) to handle non-linearly separable data. This involves transforming the input features into a higher-dimensional space.

4. Training and Evaluation:

Split the dataset into training and testing sets.

Train your SVM model on the training set using the optimization algorithm.

Evaluate the model's performance on the test set by calculating accuracy and other relevant metrics.

تمرین 4) (آقای حسین زاده) در فایل زیپ در پوشه ی PQ4 قرار گرفته است.

تمرین 5 - امتیازی) (آقای حسین زاده) در فایل زیپ در پوشه ی PQ5 قرار گرفته است.

آنچه تحویل داده میشود:

1. کد اجرایی تمرینات عملی
 2. پاسخ های تمرین تئوری در یک فایل PDF
- 📌 نکته بسیار مهم: 2 مورد بالا را zip کرده و نام آن را شماره دانشجویی خود گذاشته و فقط در کوئرا ارسال کنید.