



دانشگاه کردستان

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر و فناوری اطلاعات

عنوان درس

یادگیری ماشین

تمرین سری اول

یادگیری مفهومی و درخت‌های تصمیم

استاد

پرهام مرادی

آزاده فاروقی

تاریخ تحویل: شنبه ۲۱ فروردین

توجه : گزارش مربوط به هر بخش را در یک مستند جداگانه تحویل دهید.

در صورت مشاهده هرگونه کپی از تمرینات، نمره کل تمرین صفر منظور خواهد شد.

بخش اول : تمرینات تئوری (۴۰ نمره)

۱- الگوریتم حذف کاندید را بر روی مجموعه داده ای زیر اعمال کنید و مراحل حل را با جزییات بنویسید

Training Example	N (running nose)	C (coughing)	R (reddened skin)	Classification
d_1	+	+	+	positive (ill)
d_2	+	+	-	positive (ill)
d_3	+	-	+	negative (healthy)
d_4	-	+	+	negative (healthy)
d_5	-	-	+	negative (healthy)
d_6	-	-	-	negative (healthy)

۲- الگوریتم ID3 را بر روی مجموعه داده ای زیر اجرا کنید و به سوالات زیر پاسخ دهید.

الف: آنتروپی متغیر Edible را به دست آورید.

ب: کدام متغیر در ریشه قرار می گیرد؟

ج : درخت تصمیم را رسم کنید و مقادیر کلاس را برای نمونه های U ، V و W به دست آورید

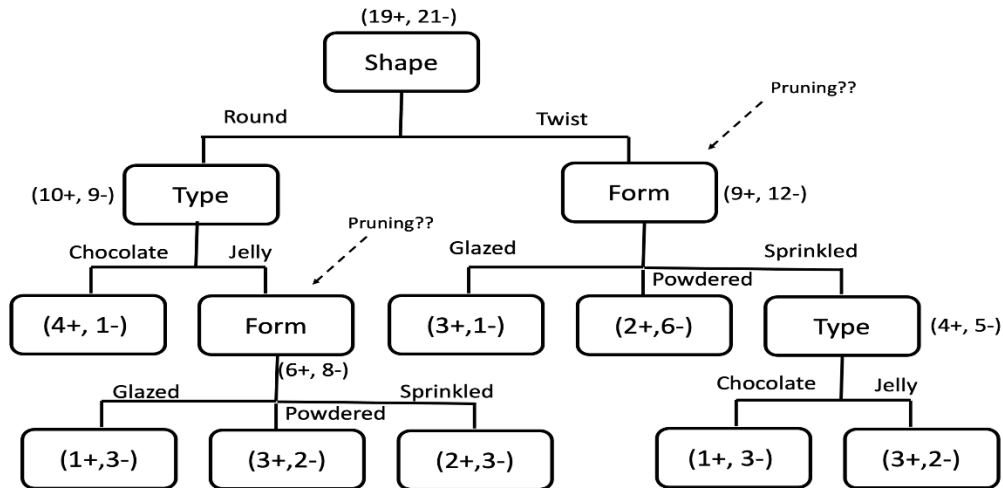
Example	NotHeavy	Smelly	Spotted	Smooth	Edible
A	1	0	0	0	1
B	1	0	1	0	1
C	0	1	0	1	1
D	0	0	0	1	0
E	1	1	1	0	0
F	1	0	1	1	0
G	1	0	0	1	0
H	0	1	0	0	0
U	0	1	1	1	?
V	1	1	0	1	?
W	1	1	0	0	?

۳- الف : آیا الگوریتم ID3 برای داده های با مقادیر نامشخص (Missing Value) کار میکند؟ روش پیشنهادی شما برای رفع این مشکل چیست؟ ب: آیا الگوریتم ID3 برای داده های با مقادیر پیوسته (Continuous Value) کار میکند؟ روش پیشنهادی شما برای رفع این مشکل چیست؟

۴- دلیل هرس کردن درخت تصمیم چیست و دو روش مختلف برای هرس کردن درخت را توضیح دهید؟

۵- در درخت ایجاد شده در شکل زیر ویژگی Form در دو گره در زیر شاخه های متفاوت بکار برده شده است. توضیح دهید هرس (pruning) کدام یک از این دو گره موجب کارایی بهتر درخت تصمیم می گردد. نکته: در درخت ایجاد شده و درخت هرس شده برچسب هر برگ با توجه به کلاس با بیشترین تکرار مشخص می گردد.

Shape	form	type	Class
Round	Glazed	Jelly	+
Round	Sprinkled	Chocolate	+
Twist	Glazed	Chocolate	-
Round	Powdered	Jelly	-
Twist	Powdered	Jelly	+
Twist	Sprinkled	Jelly	-



۶- الف : تفاوت درخت طبقه بندی (Classification Tree) و درخت رگرسیون (Regression Tree) چیست؟ شبه کد درخت رگرسیون را بنویسید.

۷- روش Cross Fold Validation را برای ارزیابی روش های طبقه بندی توضیح دهید.

بخش دوم : تمرینات عملی. درخت تصمیم برای طبقه بندی (۴۰)

هدف از این تمرین پیاده سازی درخت تصمیم بدون استفاده از کتابخانه های آماده است.

ابتدا مجموعه داده ای های زیر را از سایت UCI دانلود کنید

مجموعه داده Cancer

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

مجموعه داده Adult

<https://archive.ics.uci.edu/ml/datasets/Adult>

۱. در صورتیکه این مجموعه داده ها دارای مقادیر نامشخص هستند (Missing Value) ، استراتژی های خود را برای رفع این مشکل بیان کنید. با یکی از این استراتژی ها مقادیر نامشخص را در مجموعه داده ای تکمیل کنید
۲. در صورتیکه این مجموعه داده ها دارای ویژگی از نوع Categorical هستند با یک استراتژی مشخص آنها را به مقادیر عددی تبدیل کنید.
۳. در صورتیکه این مجموعه داده ها دارای ویژگی های با مقادیر پیوسته هستند، همه این مقادیر را در بازه ۰ و ۱ نرمال کنید
۴. مجموعه داده ای را به دو مجموعه train و test تقسیم کنید. ۷۰ درصد را به صورت تصادفی برای train و مابقی را برای تست در نظر بگیرید.
۵. درخت تصمیم را برای این دو مجموعه داده آموزش دهید (از هر نوع معیار انتخاب ویژگی می توانید استفاده کنید).
۶. ویژگی های مجموعه داده ها را به ترتیب اهمیت در ایجاد مدل مشخص نمایید.
۷. مقادیر Accuracy ، Recall ، Precision و AUC را برای درخت بدست آمده به دست بیاورید.

بخش سوم : تمرینات عملی (استفاده از کتابخانه‌های آماده مانند Sikit-learn) (۳۰)

- ۱- سوالات بخش دوم را بر روی همان مجموعه داده‌ها (Cancer و Adult) با استفاده از توابع کتابخانه‌ای آماده تکرار کنید.
- ۲- بهترین عمق درخت را با استفاده از Cross Fold Validation پیدا کنید.
- ۳- با استفاده از کتابخانه مناسب در پایتون درخت تصمیم به دست آمده را رسم کنید.
- ۴- مجموعه داده‌های زیر که از نوع رگرسیون هستند را از سایت UCI دانلود کنید و مراحل قبل را با استفاده از توابع کتابخانه‌ای آماده بر روی این مجموعه داده‌ها انجام دهید.

<https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>

<https://archive.ics.uci.edu/ml/datasets/Auto+MPG>

خروجی های مورد انتظار

- ۱- فایل Word و PDF را تحویل دهید. نام فایلها با قالب زیر باشد
YourName_YourIDNumber_Ex1.PDF (or docx)
- ۲- خروجی هر بخش را در یک فایل جداگانه تحویل دهید
- ۳- کدهای پایتون را هم در فایل توضیح دهید
- ۴- کدهای پایتون را به صورت جداگانه هم تحویل دهید