

به نام خدا



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

درس یادگیری ماشین  
استاد ناظر فرد

تمرین دوم

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

## بخش اول: پرسش‌های تشریحی

### سوال ۱

الف) اگر مقدار  $k = 1$  باشد، وقتی که داده validation هر یک از چهار داده کلاس + باشد، مدل به اشتباه آن را - پیش‌بینی می‌کند چراکه داده - در میان آن‌ها فاصله کمتری تا هر چهار داده دیگر نسبت به همدیگر دارد ( فاصله مرکز مربع تا رئوس کمتر از فاصله دو راس مربع است. بدهی است که خود این داده منفی هم به اشتباه دسته‌بندی خواهد شد. داده‌ها منفی طرف دیگر به وضوح درست دسته‌بندی می‌شوند.

باتوجه به اینکه  $k$  نباید مقدار زوج داشته باشد، حال باید  $k = 3$  را بررسی کنیم. در این حالت اتفاقی که برای چهار داده + رخ داد دیگر اتفاق نمی‌افتد چون به هر داده + دو داده + و یک داده - نزدیک است اما همچنان داده - میان آن‌ها به اشتباه دسته‌بندی می‌شود. داده‌های منفی طرف دیگر به وضوح درست دسته‌بندی می‌شوند.

اگر  $k = 5$  باشد، برای هر یک از داده‌های + همچنان سه داده + نزدیک‌تر وجود دارد که در نتیجه این‌ها به درستی ارزیابی می‌شوند. برای داده‌های - در طرف چپ هم به وضوح نتیجه دسته‌بندی درست خواهد بود. اما همچنان داده - میان داده‌های مثبت به چهار داده + نزدیک است و دسته‌بندی اشتباهی خواهد داشت.

اگر  $k \geq 7$  باشد در این صورت برای هر یک از داده‌های + تنها سه داده + باقی می‌ماند و در نتیجه همگی به اشتباه به کلاس - تعلق می‌گیرند. در نتیجه متوجه می‌شویم برای این مسئله حالت  $k = 3$  و  $k = 5$  با یک خطا بهترین حالت است.

ب) مانند تعیین مقدار هر ابرپارامتر دیگر می‌توان از یک مجموعه validation برای تعیین آن استفاده کرد. یعنی آنکه به ازای مقادیر  $k$  مختلف بررسی می‌کنیم که خطای مجموعه validation به چه شکلی بوده است. جایی که کمترین خطا را برای این مجموعه داشتیم احتمالاً  $k$  مناسب داشته است.

## سوال ۲

الگوریتم KNN را می‌توان یک الگوریتم تمایزگر دانست؛ چراکه در این الگوریتم حساب می‌شود که یک داده به کدام کلاس تعلق دارد و مرزهای کلاس‌ها مختلف تشکیل می‌شود (مانند سلول‌های voronoi برای 1NN) و نمی‌توان الگوی یک کلاس را محاسبه کرد و توسط آن داده‌های جدید ایجاد کرد.

الگوریتم درخت تصمیم را هم باید یک الگوریتم تمایزگر دانست؛ چراکه در این الگوریتم هم باز صرفاً بررسی می‌شود که یک داده به کدام کلاس تعلق دارد و می‌توان مرزها را برای کلاس‌های مختلف تشکیل داد ولی نمی‌توان توزیع هر کلاس را بدست آورد.<sup>۱</sup>

## سوال ۳

(الف)

$$\sigma(a) = \frac{1}{1 + e^{-a}} \rightarrow \frac{d\sigma(a)}{da} = \frac{e^{-a}}{(1 + e^{-a})^2} = \sigma(a) * \left( \frac{e^{-a}}{1 + e^{-a}} \right) = \sigma(a) * \left( 1 - \frac{1}{1 + e^{-a}} \right) \\ = \sigma(a) * (1 - \sigma(a))$$

(ب)

$$\hat{y}^{(i)} = p(C_1|x^{(i)}) = \sigma(w^T x^{(i)})$$

$$f = -\log \sigma(w^T x)$$

(ج)

$$\frac{\partial f}{\partial w} =$$

(د)؟

<sup>1</sup> <https://stats.stackexchange.com/questions/105979/is-knn-a-discriminative-learning-algorithm>

<sup>2</sup> <https://stats.stackexchange.com/questions/12421/generative-vs-discriminative>

## سوال ٤

(الف)

$$p(\text{buy} = \text{yes}) = \frac{9}{14}, p(\text{buy} = \text{no}) = \frac{5}{14}$$

---

$$p(\text{buy} = \text{yes} | X_1) = \frac{p(X_1 | \text{buy} = \text{yes}) * p(\text{buy} = \text{yes})}{p(X_1)}$$

$$p(X_1 | \text{buy} = \text{yes})$$

$$\begin{aligned} &= p(\text{age} = \text{youth} | \text{buy} = \text{yes}) * p(\text{income} = \text{high} | \text{buy} = \text{yes}) \\ &* p(\text{student} = \text{yes} | \text{buy} = \text{yes}) * p(\text{credit} = \text{fair} | \text{buy} = \text{yes}) \\ &= \frac{2}{9} * \frac{2}{9} * \frac{6}{9} * \frac{6}{9} \end{aligned}$$

$$p(\text{buy} = \text{no} | X_1) = \frac{p(X_1 | \text{buy} = \text{no}) * p(\text{buy} = \text{no})}{p(X_1)}$$

$$p(X_1 | \text{buy} = \text{no})$$

$$\begin{aligned} &= p(\text{age} = \text{youth} | \text{buy} = \text{no}) * p(\text{income} = \text{high} | \text{buy} = \text{no}) \\ &* p(\text{student} = \text{yes} | \text{buy} = \text{no}) * p(\text{credit} = \text{fair} | \text{buy} = \text{no}) \\ &= \frac{3}{5} * \frac{2}{5} * \frac{1}{5} * \frac{2}{5} \end{aligned}$$

$$p(\text{buy} = \text{yes} | X_1) \approx \frac{0.014}{p(X_1)}, p(\text{buy} = \text{no} | X_1) \approx \frac{0.006}{p(X_1)}$$

$$\rightarrow p(\text{buy} = \text{yes} | X_1) > p(\text{buy} = \text{no} | X_1)$$

---

$$p(\text{buy} = \text{yes} | X_2) = \frac{p(X_2 | \text{buy} = \text{yes}) * p(\text{buy} = \text{yes})}{p(X_2)}$$

$$(X_2 | \text{buy} = \text{yes})$$

$$\begin{aligned} &= p(\text{age} = \text{senior} | \text{buy} = \text{yes}) * p(\text{income} = \text{low} | \text{buy} = \text{yes}) \\ &* p(\text{student} = \text{no} | \text{buy} = \text{yes}) * p(\text{credit} = \text{excellent} | \text{buy} = \text{yes}) \\ &= \frac{3}{9} * \frac{3}{9} * \frac{3}{9} * \frac{3}{9} \end{aligned}$$

$$p(\text{buy} = \text{no} | X_2) = \frac{p(X_2 | \text{buy} = \text{no}) * p(\text{buy} = \text{no})}{p(X_2)}$$

$$(X_2 | \text{buy} = \text{no})$$

$$= p(\text{age} = \text{senior} | \text{buy} = \text{no}) * p(\text{income} = \text{low} | \text{buy} = \text{no}) \\ * p(\text{student} = \text{no} | \text{buy} = \text{no}) * p(\text{credit} = \text{excellent} | \text{buy} = \text{no}) \\ = \frac{2}{5} * \frac{1}{5} * \frac{4}{5} * \frac{3}{5}$$

$$p(\text{buy} = \text{yes} | X_2) \approx \frac{0.007}{p(X_2)}, p(\text{buy} = \text{no} | X_2) \approx \frac{0.013}{p(X_2)}$$

$$\rightarrow p(\text{buy} = \text{yes} | X_2) < p(\text{buy} = \text{no} | X_2)$$

$$p(\text{buy} = \text{yes} | X_3) = \frac{p(X_3 | \text{buy} = \text{yes}) * p(\text{buy} = \text{yes})}{p(X_3)}$$

$$(X_3 | \text{buy} = \text{yes})$$

$$= p(\text{age} = \text{middle} | \text{buy} = \text{yes}) * p(\text{income} = \text{medium} | \text{buy} = \text{yes}) \\ * p(\text{student} = \text{no} | \text{buy} = \text{yes}) * p(\text{credit} = \text{fair} | \text{buy} = \text{yes}) \\ = \frac{4}{9} * \frac{4}{9} * \frac{3}{9} * \frac{6}{9}$$

$$p(\text{buy} = \text{no} | X_3) = \frac{p(X_3 | \text{buy} = \text{no}) * p(\text{buy} = \text{no})}{p(X_3)}$$

$$(X_3 | \text{buy} = \text{no})$$

$$= p(\text{age} = \text{middle} | \text{buy} = \text{no}) * p(\text{income} = \text{medium} | \text{buy} = \text{no}) \\ * p(\text{student} = \text{no} | \text{buy} = \text{no}) * p(\text{credit} = \text{fair} | \text{buy} = \text{no}) = 0$$

$$p(\text{buy} = \text{yes} | X_3) > p(\text{buy} = \text{no} | X_3)$$

(ب)

ابتدا باید بررسی کرد که برای ریشه درخت کدام ویژگی مناسبتر است.

$$E(\text{age} = \text{youth}) = -\frac{3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} \approx 0.442 + 0.528 = 0.97$$

$$E(\text{age} = \text{middle}) = -1 \log 1 = 0$$

$$E(\text{age} = \text{senior}) = -\frac{3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} \approx 0.442 + 0.528 = 0.97$$

$$Gain(age) = E(S) - \left( \frac{5}{14} * 0.97 + \frac{5}{14} 0.97 + \frac{4}{14} * 0 \right) = E(s) - 0.692$$

$$E(income = high) = -\frac{2}{4} \log \frac{2}{4} - \frac{2}{4} \log \frac{2}{4} = 1$$

$$E(income = medium) = -\frac{4}{6} \log \frac{4}{6} - \frac{2}{6} \log \frac{2}{6} \approx 0.389 + 0.528 = 0.917$$

$$E(income = low) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \approx 0.311 + 0.5 = 0.811$$

$$Gain(income) = E(S) - \left( \frac{4}{14} * 1 + \frac{6}{14} * 0.917 + \frac{4}{14} * 0.811 \right) = E(s) - 0.910$$

$$E(student = yes) = -\frac{6}{7} \log \frac{6}{7} - \frac{1}{7} \log \frac{1}{7} \approx 0.190 + 0.401 = 0.591$$

$$E(student = no) = -\frac{3}{7} \log \frac{3}{7} - \frac{4}{7} \log \frac{4}{7} \approx 0.523 + 0.461 = 0.984$$

$$Gain(Student) = E(S) - \left( \frac{7}{14} * 0.591 + \frac{7}{14} * 0.984 \right) = E(S) - 0.787$$

$$E(credit = fair) = -\frac{2}{8} \log \frac{2}{8} - \frac{6}{8} \log \frac{6}{8} \approx 0.5 + 0.311 = 0.811$$

$$E(credit = excellent) = -\frac{3}{6} \log \frac{3}{6} - \frac{3}{6} \log \frac{3}{6} = 1$$

$$Gain(credit) = E(S) - \left( \frac{8}{14} * 0.811 + \frac{6}{14} * 1 \right) = E(s) - 0.892$$

به نظر می‌رسد برای ریشه پرسش راجع به age بهترین گزینه است.

---

حال باید بررسی کرد که برای شاخه age = youth کدام ویژگی مناسب‌تر است.

| age   | income | student | credit    | Buy |
|-------|--------|---------|-----------|-----|
| Youth | High   | No      | Fair      | NO  |
| Youth | High   | No      | Excellent | NO  |
| Youth | Medium | No      | Fair      | NO  |
| Youth | Low    | Yes     | Fair      | YES |
| Youth | Medium | Yes     | Excellent | YES |

طبیعتاً برای این شاخه هم می‌توان مشابه با محاسبات گره ریشه پیش رفت؛ اما پیش از آن توجه کنید که اگر student مورد سوال قرار بگیرد قطعا دو گره خالص ایجاد خواهد شد این درحالی است که در مابقی ویژگی‌ها حداقل یک گره ناخالص باقی می‌ماند. ( برای income، مقدار medium؛ برای credit، مقدار fair) پس قطعا Gain ویژگی Student برابر بیشینه مقدار که آنتروپی همین گره age = youth باشد است که از مابقی بیشتر است و نیازی به محاسبات بیشتر نیست.

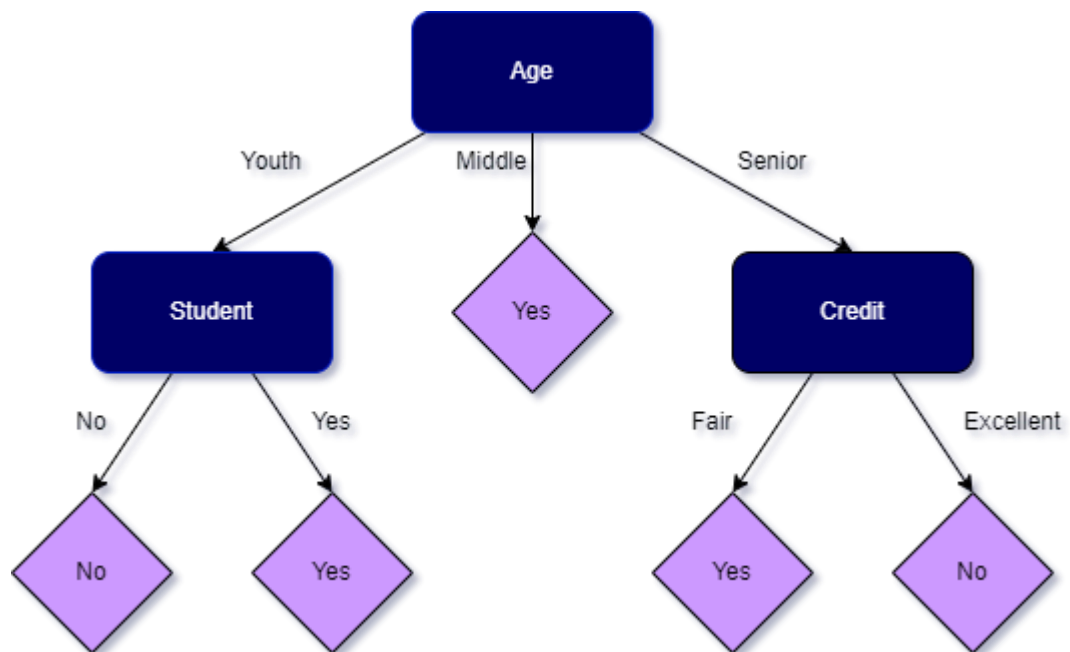
برای شاخه age = middle یک گره خالص باقی می‌ماند که نیاز به ادامه ندارد و می‌دانیم به پاسخ yes خواهیم رسید.

| age    | income | student | credit    | Buy |
|--------|--------|---------|-----------|-----|
| Middle | High   | No      | Fair      | YES |
| Middle | Low    | Yes     | Excellent | YES |
| Middle | Medium | No      | Excellent | YES |
| Middle | High   | Yes     | Fair      | YES |

برای شاخه آخر یعنی age = senior، واضحاً و مشابه با بحث گره age = youth انتخاب ویژگی credit بهترین گزینه است چراکه دو گره خالص حاصل می‌شود و Gain این ویژگی برابر با آنتروپی گره age = senior است که بیشترین Gain ممکن است. برای دو ویژگی دیگر هم مشخص است که گره‌های حاصل تماماً خالص نخواهند بود.

| age    | income | student | credit    | Buy |
|--------|--------|---------|-----------|-----|
| Senior | Medium | No      | Fair      | YES |
| Senior | Low    | Yes     | Fair      | YES |
| Senior | Low    | Yes     | Excellent | NO  |
| Senior | Medium | Yes     | Fair      | YES |
| Senior | Medium | No      | Excellent | NO  |

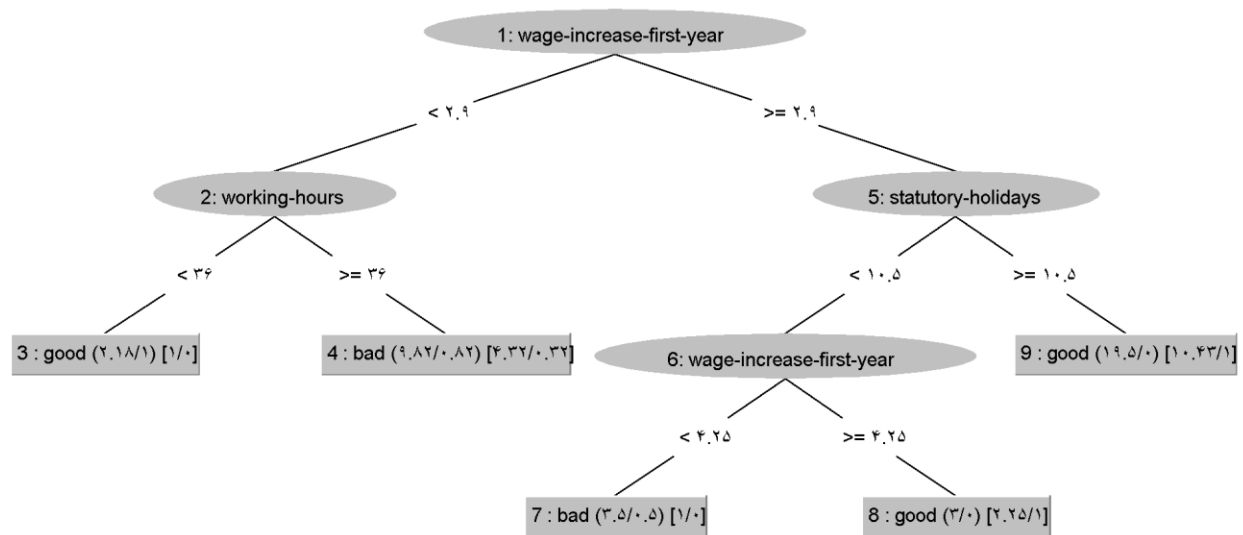
درخت نهایی برابر خواهد بود با:



## سوال ۵

الف) برای این سوال از الگوریتم REPTree موجود در نرم افزار استفاده کردم و سایر پارامترها را تغییر ندادم.





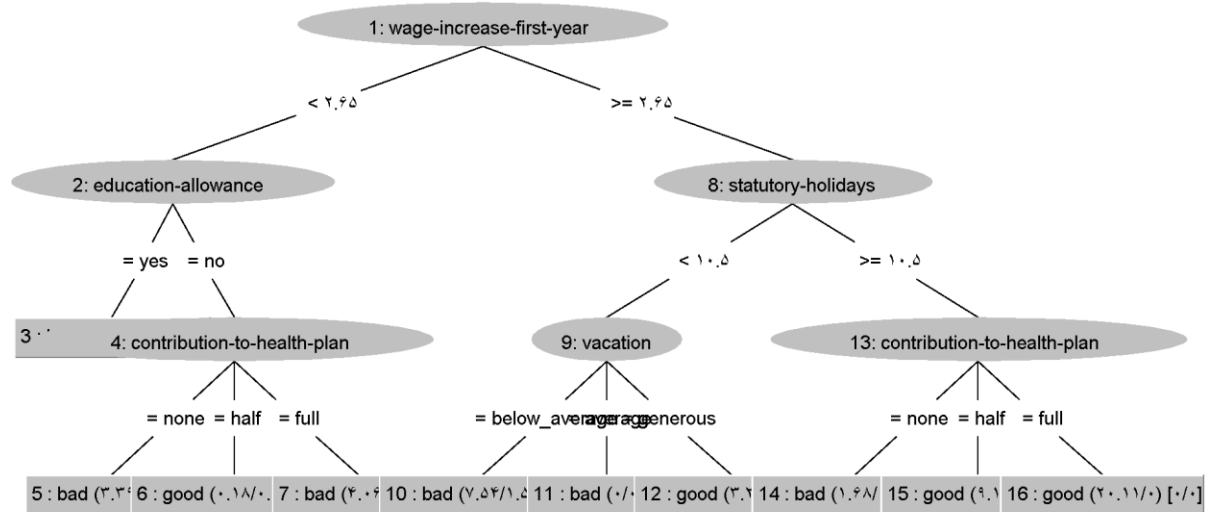
(ب)

=== Confusion Matrix ===

```

a  b  <-- classified as
14  6  |  a = bad
 6 31  |  b = good
  
```

(ج)



د) مطابق انتظار درخت هرس نشده تعداد گره بیشتری دارد. همچنین می‌توان مشاهده کرد که درصد خلوص در گره‌های انتهایی درخت هرس نشده بیشتر است.