تمرین چهارم ML

امیرحسین باقری 98105621

سوال یک

الف

مطابق با فرض مسئله داده ها خطی جدایی پذیرند. بنابراین w موجود است که داده ها را جدا کند. حال در بین تمام w های ممکن کم‌ترین اندازه را در نظر می‌گیریم.

$B = \min \|w\| \; : \forall i \in M, \; y_i \langle w, x_i \rangle \geq 1$

مطابق الف داریم:

$R = \max_i \|X_i\|_2$

$w_B \rightarrow$ minimum of all $|w|$ that we can apply

حال داریم

ابتدا یک $V = \frac{|w_B|}{B}$ را در نظر بگیرید؛ $\Rightarrow V$ is unit vector $\quad |V| = 1 \longleftarrow$ ابتدا روابط بالا را بازنویسی می‌کنیم.

$V = \frac{|w_B|}{B} : \forall i \in M \quad y_i \langle V, x^{(i)} \rangle \geq \frac{1}{B} = \gamma$

$R = \max \|X_i\|$

بنابراین یک unit_vector : V وجود دارد. به طوری که

for all $x_i \rightarrow y^{(i)} (x^{(i)})^T V \geq \gamma$

حال ثابت می‌کنیم که حداکثر تعداد اشتباهاتی که الگوریتم Perceptron مرتکب می‌شود حداکثر برابر $\left(\frac{D}{\gamma}\right)^2$ است.

Proof:

الگوریتم Perceptron تنها وقتی missclassification ها اعمال می‌شود حال فرض کنید که بردار w^k بردار جداکننده مورد استفاده در k امین اشتباه در داده (x^i, y^i) است. داریم $[w^1 = [0]]$

$w^{k+1} = w^k + y^{(i)} x^{(i)} \quad , \quad (x^i)^T w^k (y^i) \leq 0$

$(w^{k+1})^T V = w^{kT} V + \underbrace{y^{(i)} (x^{(i)})^T V}_{\geq \gamma} \geq (w^k)^T V + \gamma$

Scanned by CamScanner

بنابراین بنابر استقرا داریم:

$$w^{1^T} = 0$$

$$(w^2)^T u \geqslant Y \quad \rightarrow$$ پایه استقرا

$$(w^3)^T u \geqslant (w^2)^T u + Y \geqslant 2Y$$

$$(w^{(k+1)})^T u \geqslant (w^k)^T u + Y \geqslant kY \rightarrow$$ بنابر استقرا

$$(w^k)^T u \geqslant (w^{k-1})^T u + Y \geqslant (k-1)Y$$

$$(w^{k+1})^T u \geqslant (w^k)^T u + Y \geqslant (k-1)Y + Y \geqslant kY$$

$$\Longrightarrow (w^{k+1})^T u \geqslant kY \quad \blacksquare$$

از طرفی داریم:

$$\|w^{k+1}\|^2 = \|w^k + y^{(i)} x^{(i)}\|^2 = \|w^k\|^2 + |x^i|^2 + \underbrace{2 y^{(i)} (x^{(i)})^T w^k}_{\leqslant 0} \rightarrow$$ چون missclassify تنها محاسبه می‌شود.

$$\Longrightarrow \|w^{k+1}\|^2 \leqslant \|w^k\|^2 + \|x^i\|^2$$

$$\|x^i\|^2 \leqslant R^2$$

$$\Longrightarrow \|w^{k+1}\|^2 \leqslant \|w^k\|^2 + R^2$$

حال داریم:

$$w^1 = 0$$

$$\|w^2\|^2 \leqslant R^2 \rightarrow$$ پایه استقرا ~~$w^T u$~~ ~~که میتوان دید است unit vector~~

$$\|w^k\|^2 \leqslant (k-1)R^2$$ فرض

$$\|w^{k+1}\|^2 \leqslant \|w^k\|^2 + R^2 \leqslant (k-1)R^2 + R^2 \leqslant kR^2 \rightarrow$$ حکم استقرا $\blacksquare$

$$\|w^{k+1}\|^2 \leqslant kR^2 \Longrightarrow \sqrt{k}R \geqslant \|w^{k+1}\| \geqslant \underline{(w^{k+1})^T u}$$

I $\sqrt{\sum w_i^2} \geqslant \sum a_i w_i$ s.to $\sqrt{\sum a_i^2} = 1$. چون u یک بردار با طول واحد است، ⟸ کی حل است . اشتباه سوال توضیحات

$$\text{I} \rightarrow \sqrt{k}R \geqslant (w^{k+1})^T u \geqslant kY \Longrightarrow \sqrt{k} \leqslant R/Y$$

استفاده از نامساوی I

$$\Longrightarrow k \leqslant \frac{R^2}{Y^2}$$

ادامه سوال یک

الف ←

بنابراین داریم $K \leq \left(\frac{R}{\gamma}\right)^2$ بنابراین حداکثر تعداد اشتباهات k برابر $\left(\frac{R}{\gamma}\right)^2$ است.

$$\gamma = \frac{1}{B} \Rightarrow K \leq (RB)^2$$

بررسی بیشتر نامساوی (I)

$$(w^{k+1})^T v = \|w^{k+1}\| \|v\| \cos\phi \leq \|w^{k+1}\|$$

بنابراین از این نامساوی در قسمت قبل استفاده کردیم.

ب

تمام فروض بخش قبل را در نظر بگیرید ←

there is a v with $\|v\|=1$ and for all x^i, y^i $(x^i)^T v (y^i) \geq \gamma$

← از آنجا که استراها مانند بخش قبل است از اثبات دوباره آنها خودداری می‌کنیم.

$$w^{k+1} = w^k + \eta x^{(i)} y^{(i)}$$

$$\Rightarrow (w^{k+1})^T v = (w^k)^T v + \eta \underbrace{(x^i)^T y^i v}_{\geq \gamma} \geq (w^k)^T v + \eta \gamma$$

$$\xRightarrow{\text{بنابر استرا}} (w^{k+1})^T v \geq k \eta \gamma \rightarrow (1)$$ نامساوی اول:

$$\|w^{k+1}\|^2 = \|w^k\|^2 + \|\eta^2 x^i y^i\|^2 + \underbrace{2\eta \|x^{iT} w^k\|}_{\leq 0} \leq \|w^k\|^2 + \eta^2 \|x^i\|^2$$

$$\xRightarrow{\text{بنابر استرا}} \|w^{k+1}\|^2 \leq k\eta^2 \|x^i\|^2 \leq k\eta^2 R^2 \rightarrow$$ نامساوی دوم

$$(w^{k+1})^T v \leq \sqrt{k}\, \eta R \Rightarrow k\eta\gamma \leq \sqrt{k}\, \eta R \Rightarrow K \leq \left(\frac{R}{\gamma}\right)^2$$

$$\Rightarrow K \leq (RB)^2 \quad (1) \rightarrow$$ مانند قسمت الف

سوال ۳

$R(h) = P(Y \neq h(x))$

$$P(Y \neq h(x) \mid x) = P(Y=1, h(x)=0 \mid x) + P(Y=0, h(x)=1 \mid x)$$

$$P(Y=1 \mid x)(h(x)==0) + P(Y=0 \mid x)(h(x)==1) = \quad \leftarrow$$ اگر x را داشته باشیم h(x) را داریم

$$P(Y=1 \mid x)(1 - h(x)==1) + (1-P(Y=1 \mid x))(h(x)==1) =$$

$$P(Y=1 \mid x) + (h(x)==1)[1-2P(Y=1 \mid x)] = P(Y=1 \mid x) + I(h(x)=1)[1-2P(Y=1 \mid x)]$$

*۲ $\forall x,h :\ R(h) - h(h^*) \geq 0$

* یا معادلاً : $E(h-h^*) \geq 0$ ← برای اینکه ثابت کنیم که $h^*(x)$ ، $R(h)$ را کمینه می‌کند کافی است ثابت کنیم که

$$R(h) - R(h^*) = \cancel{P(Y=1 \mid x)} + I(h(x)=1)[1-2P(Y=1 \mid x)] - \cancel{P(Y=1 \mid x)} - I(h^*(x)=1)[1-2P(Y=1 \mid x)]$$

$$R(h) - R(h^*) = \underbrace{[I(h(x)=1) - I(h^*(x)=1)]}_{\alpha}\underbrace{[1-2P(Y=1 \mid x)]}_{\beta}$$

$$\alpha = I(h(x)=1) - \overset{1}{\cancel{I(h^*(x)=1)}} = I(h(x)=1) - 1 \leq 0 \quad \leftarrow P(Y=1 \mid x) > 1/2$$ حال فرض کنید که

$$\beta = 1 - 2 \times (>1/2) < 0 \quad \Rightarrow \quad \alpha\beta \geq 0 \checkmark$$

$$\alpha = I(h(x)=1) - \overset{0}{\cancel{I(h^*(x)=1)}} = I(h(x)=1) \geq 0 \quad \leftarrow$$ باشد $P(Y=1 \mid x) \leq 1/2$ حال فرض کنید که

$$\beta = 1 - 2P(Y=1 \mid x) \geq 0 \quad \Rightarrow \quad \alpha\beta \geq 0$$

$$\forall h,x :\ R(h(x)) - R(h^*(x)) \geq 0$$ داریم $\Longleftarrow$

حال باید امید ریاضی بگیریم: برای تمام x ها:

$$E(h-h^*) = \int_{R_x} P(x)\, \overbrace{\alpha(x)\beta(x)}^{\geq 0} \geq 0 \Rightarrow$$ h^* پاسخ بهینه است.

بخش ب صفحه بعد

ب

$$R(h)-R(h^*) = E(h-h^*)_x = \int_x \alpha_{(x)}\beta_{(x)}P_{(x)}\,dx$$

داریم

چون حاصل ضرب $\alpha\beta \geq 0$ است $\Leftarrow$

$$\left.\begin{array}{l}|\alpha\beta| \geq 0\\ \alpha\beta \geq 0\end{array}\right\} \Rightarrow |\alpha\beta| = |\alpha||\beta|$$

$\Leftarrow$

$$\int_x \left|I(h(x)=1)-I(h^*(x)=1)\right| \left|1-2P(Y=1|x)\right| P(x)\,dx =$$

$0 \Leftarrow I(h(x)=1) = I(h^*(x)=1)$ اگر

$1 \Leftarrow I(h(x)=1) \neq I(h^*(x)=1)$

$$\left|I(h(x)\neq h^*(x))\right|$$

$$= \int_x \left|I(h(x)\neq h^*(x))\right| \left|1-2P(Y=1|x)\right| P(x)\,dx$$

~~$P(Y=1|x)\neq 1/2$~~

حال داریم: $\leftarrow$

$$\hat{m}(x) > \frac{1}{2},\ m^*(x) \leq \frac{1}{2} \Rightarrow \hat{m}(x)-m^*(x) \geq \hat{m}(x)-1/2$$

$$\hat{m}(x) \leq \frac{1}{2},\ m^*(x) > 1/2 \Rightarrow m^*(x)-\hat{m}(x) \geq m^*(x)-1/2 \geq 1/2-\hat{m}(x)$$

$$\Rightarrow |\hat{m}(x)-m^*(x)| \geq |1/2-\hat{m}(x)| = \left|\frac{1}{2}-P(Y=1|x)\right| \xrightarrow{2\times}$$

$$2|\hat{m}(x)-m^*(x)| \geq |1-2P(Y=1|x)| = \beta \qquad \text{I}$$

$$R(\hat{h})-R(h^*) = \int_x \overbrace{I(h(x)\neq h^*(x))}^{\text{یا ۱ یا صفر}} |1-2P(Y=1|x)|P(x)\,dx \Rightarrow$$

حال داریم $\leftarrow$

$$R(h)-R(h^*) = \int_x |1-2P(Y=1|x)|P(x)\,dx \overset{\text{I}}{\leq} \int_x 2|\hat{m}(x)-m^*(x)|P(x)\,dx$$

$$\Rightarrow R(h)-R(h^*) \leq 2\int_x |\hat{m}(x)-m^*(x)|P(x)\,dx$$

سوال ۴

الف)

$\min H(Y|X) = \min -\sum_i P(x=i)\left[\sum_j P(y=j|x=i)\log P(y=j|x=i)\right]$

ابتدا دسته‌بندی ریشه را می‌یابیم:

نوع غذا →
- فست فود ۳ → ۲
- ایرانی ۳ → ۲
- آسیایی ۵ → ۲

$$\to -\left(\frac{3}{11}\times\left(\frac{2}{3}\times\log\frac{2}{3}+\frac{1}{3}\times\log\frac{1}{3}\right)\right)+ -\left(\frac{3}{11}\times\left(\frac{2}{3}\times\log\frac{2}{3}+\frac{1}{3}\times\log\frac{1}{3}\right)\right)+ -\left(\frac{5}{11}\times\left(\frac{2}{5}\times\log\frac{2}{5}+\frac{3}{5}\log\frac{3}{5}\right)\right) = ۰/۹۴$$

رنج قیمت =
- ارزان ۴ → ۲
- متوسط → ۵ → ۳
- گران ۲ → ۰

→ $H(Y|X) =$ ۰/۸۰

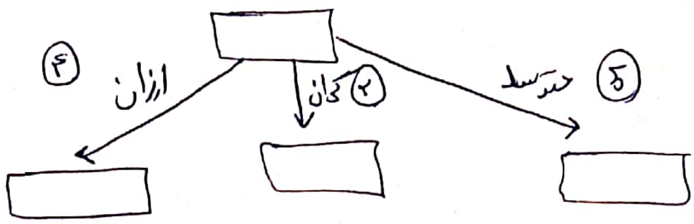
حمل =
- [illegible] ۳ → ۲
- [illegible] ۳ → ۱
- [illegible] ۴ → ۲

$H(Y|X) =$ ۰/۹۷

محدودیت =
- Vegeterian → ۴ → 2
- Glutenfree → 3 → 1
- None → ۴ → 2

$H(Y|X) =$ ۰/۹۷

$\Leftarrow \min_x H(Y|X) \to$ ۰/۸ → رنج قیمت.



ارزان →
- نوع غذا
  - فست فود → ۱ → ۰
  - ایرانی → ۱ → ۱
  - آسیایی → ۲ → ۱

  $H(Y|X) =$ ۰/۵
- حمل
  - [illegible] → ۱ → ۱
  - [illegible] → 2 → 1
  - 1 → ۰

  $H(Y|X) =$ ۰/۵
- محدودیت
  - veg → 2 → 1
  - glu → 1 → ۰
  - None → 1 → 1

  → $H(Y|X) =$ ۰/۵

حال برای هر ریشه دوباره مراحل را تکرار کنیم.

→ یکی را رندم برمی‌گزینیم:

محدودیت را انتخاب می‌کنیم.

ادامه ۴ الف)

گران → | نوع غذا → | fast food 1 → ۰ | آسیایی 1 → ۰

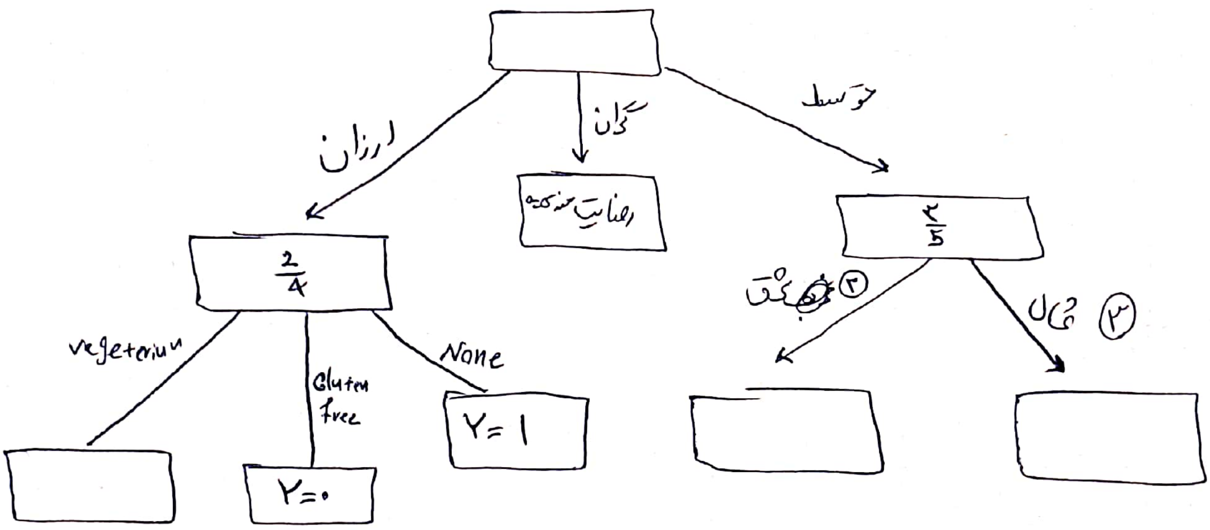
رستوران های گران $Y=0$ است ⟸ نیاز به تقسیم بندی نیست؛

گران → $Y=0$

متوسط | نوع غذا → | فست فود ۴ → ۱ | ایرانی ۲ → ۱ | آسیایی ۲ → ۱ — $H(Y|X) = .8$

حمل → | ~~...~~ | ترق ۳ → ۳ | تمال → ۳ → ۱ — $H(Y|X) = .55$ ⟹ انتخاب حمل.

محدودیت → | veg → 1 → 1 | Gluten → 2 → 1 | None → 2 → 1 — $H(Y|X) = .8$



ارزان — None → 1 → 1 → نیاز به ادامه نیست → $Y=1$ ← برای ارزان داریم

ارزان — Gluten free → ۱ → ۰ → نیاز به دسته بندی نیست → $Y=0$

ارزان — vegeterian → | نوع غذا → { فست فود → ۱ → ۰ , آسیایی → ۱ → ۱ } ⟹ $H(Y|X)=0$ → برچسب نوع غذا

حمل → { خوب → ۲ → ۱ } → $H(Y|X)=1$

متوسط — تمال → | نوع غذا = { ایرانی → ۱ → ۰ , آسیایی → ۱ → ۰ , فست فود → ۱ → ۱ } — $H(Y|X)=0$ ← حال برای ادامه دهیم

محدودیت = { None → ۲ → ۱ , gluten free → 1 → ۰ } ⟹ $H(Y|X)=1$

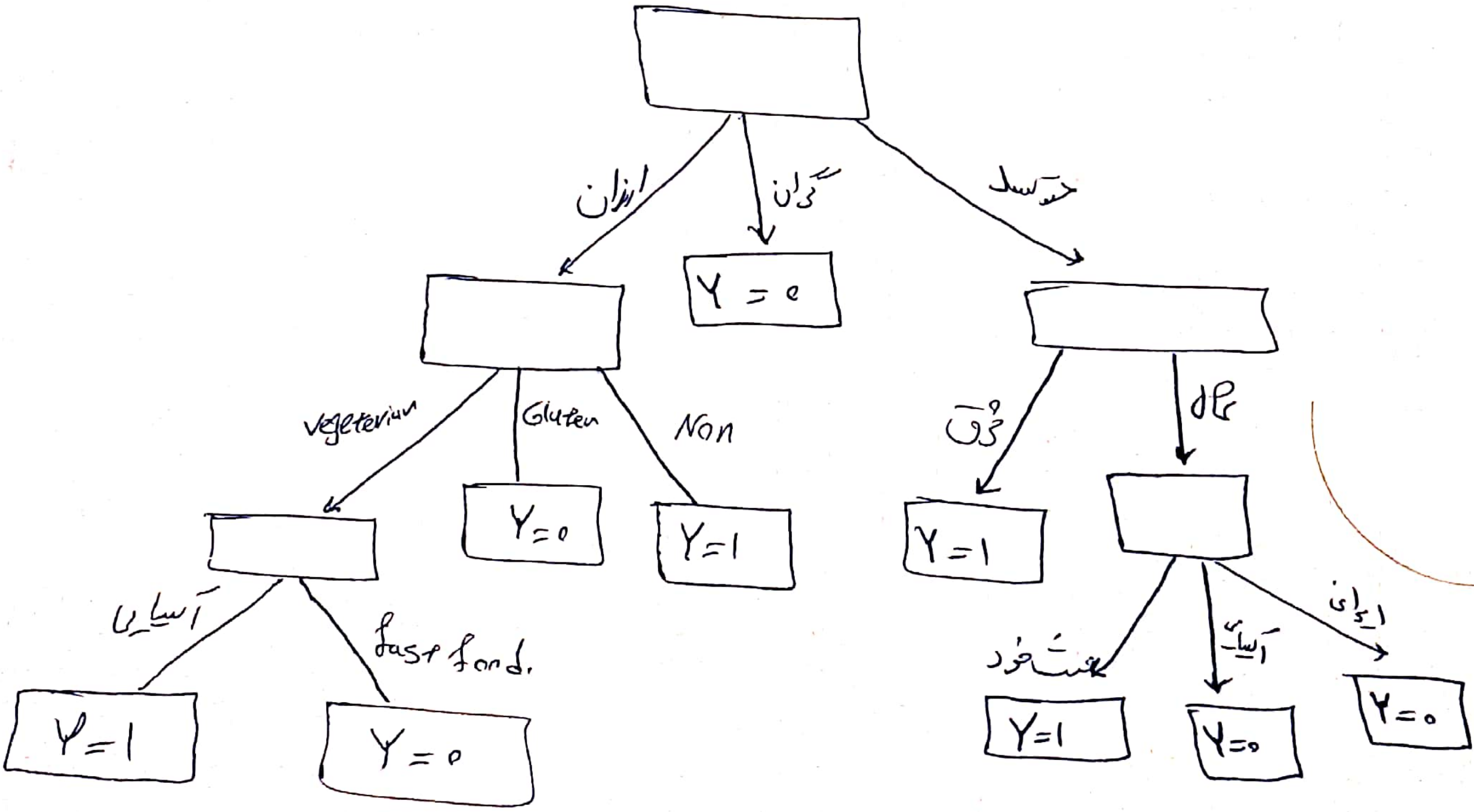
متوسط — ترق → | نوع غذا | محدودیت | رضایت مشتری 1 نمونه → نیاز به دسته بندی نیست و ۱ است.

ادامه ۱۴ الف

بنابر این دسته ارزان را در Vegeterian را برحسب نوع غذا ادامه می دهیم.

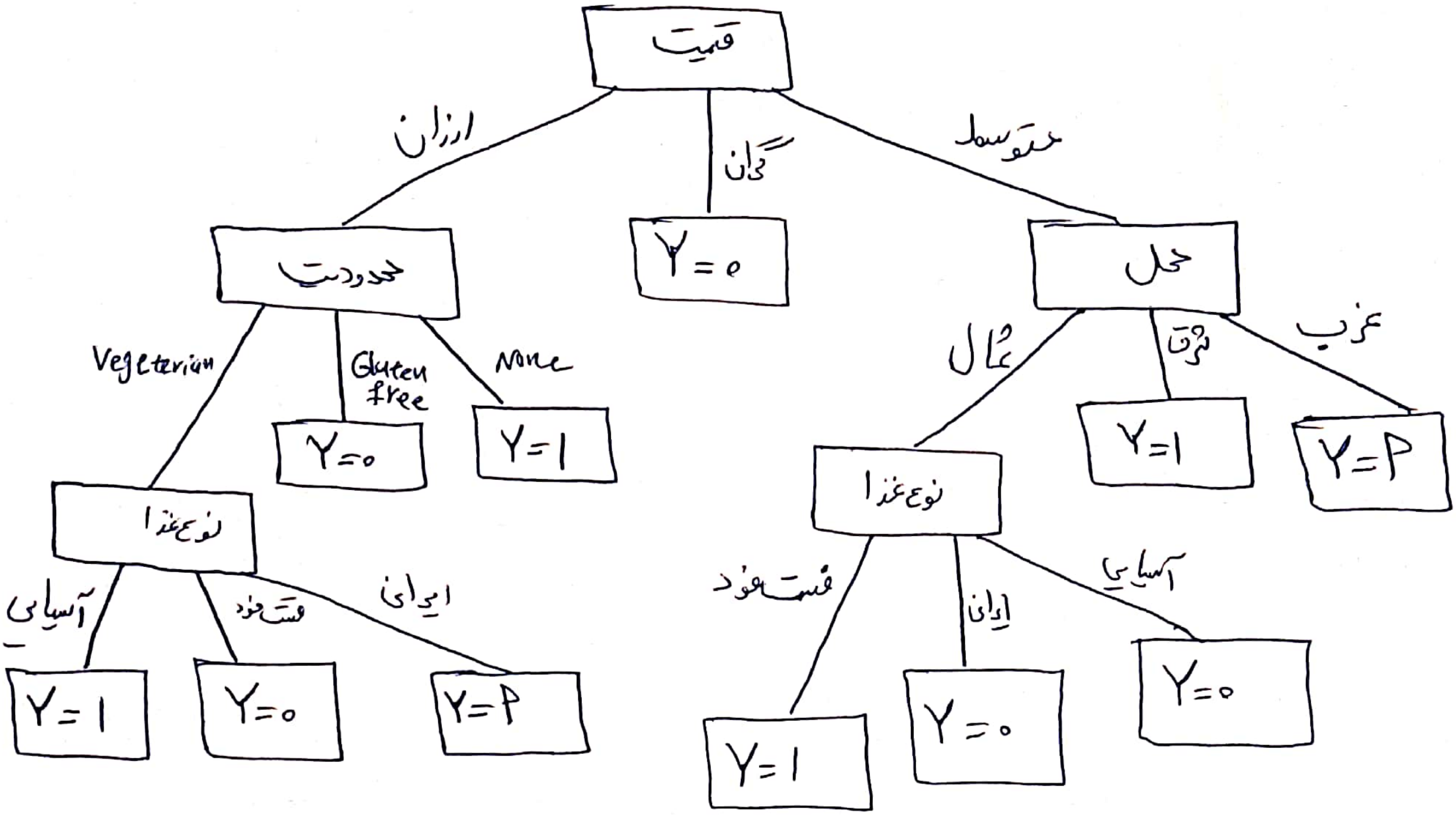
دسته متوسط شمال را نیز برحسب نوع غذا.

درخت



ادامه 4

$y=1 \longrightarrow$ class 1
$y=0 \longrightarrow$ class 0
$y=P \longrightarrow$ use a policy you want $\begin{cases} \text{random } \# \\ \text{use prior } \# \end{cases}$ ← درخت کامل به صورت زیر است



ب) Predict = output vector = [1, 1, 0, 1, 0] , original = [0, 1, 0, 1, 0]

$accuracy = \frac{4}{5} \times 100 = 80\%$

پ)

f_1 score $= 80\%$

سوال 5

الف

تابع likelihood را می‌نویسیم: (تعیین کننده کلاس هر داده را داریم)

$$P(\phi_n, t_n | \pi_k) = \prod_{n=1}^{N} P(\phi_n | C_j)\, \pi_j \qquad j = \text{Class of each } t_n \text{ that we have.}$$

حال به صورت برداری عبارت $P(\phi_n|C_k)\pi_k = \prod_{k=1}^{K} (P(\phi_n|C_k)\pi_k)^{t_{nk}}$ نوشت زیرا t_n تنها یک ۱ دارد و بقیه صفرند. بنابراین داریم:

$$P(\phi_n, t_n | \pi_k) = \prod_{n=1}^{N} \prod_{k=1}^{K} (P(\phi_n|C_k)\pi_k)^{t_{nk}}$$

$$L = \ln P(\phi, t_n | \pi_k) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\ln P(\phi_n|C_k) + \ln \pi_k \right)$$

تعیین کننده داریم $\Leftarrow \sum_{j=1}^{K} \pi_j = 1$

maximize L s.t $\sum_{j=1}^{K} \pi_j = 1$

$$\Longrightarrow L' = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\ln P(\phi_n|C_k) + \ln \pi_k \right) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \Longrightarrow$$

$$\frac{\partial L'}{\partial \pi_k} = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\cancel{\frac{\partial \ln P(\phi_n|C_k)}{\partial \pi_k}} + \frac{1}{\pi_k} \right) + \lambda = 0 \Longrightarrow$$

$$\sum_{n=1}^{N} \underbrace{\sum_{k=1}^{K} \frac{t_{nk}}{\pi_k}}_{\frac{t_{nk}}{\pi_k}} + \lambda = 0 \Longrightarrow \sum_{n=1}^{N} \frac{t_{nk}}{\pi_k} + \lambda = 0 \Longrightarrow$$

$$\Longrightarrow -\lambda \pi_k = \sum_{n=1}^{N} t_{nk} = N_k \qquad \boxed{\pi_k = \frac{N_k}{N}}$$

از طرفی

$$\sum_{k=1}^{K} N_k = N \Longrightarrow -\lambda \cancel{\sum_{k=1}^{K} \pi_k}^{\,1} = N \Longrightarrow \lambda = -N$$

ب)

از قسمت قبل Likelihood function را می‌نویسیم:

$$L = \ln P(\phi, t_n \mid \pi_k) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left[\ln P(\phi_n \mid c_k) + \ln \pi_k\right]$$

$$\cancel{L = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}[}$$

حال می‌دانیم که

$$P(\phi_n \mid c_k) \sim N(\phi \mid \mu_k, \Sigma) \Leftarrow \text{پیروی می‌کند}$$

$$\Leftarrow \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(\frac{(\phi-\mu_k)^T \Sigma^{-1} (\phi-\mu_k)}{-2}\right)$$

$$\ln P(\phi \mid c_k) = \ln\left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\right) + (\phi-\mu_k)^T \Sigma^{-1} (\phi-\mu_k)$$

$$L = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left[\ln \pi_k + \ln \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} + \frac{(\phi_n-\mu_k)^T \Sigma^{-1} (\phi-\mu_k)}{-2}\right]$$

$$\frac{\partial L}{\partial \mu_k} = \sum_{n=1}^{N} t_{nk} \Sigma^{-1} (\phi_n - \mu_k) = 0 \xrightarrow{\times \Sigma} \sum_{n=1}^{N} t_{nk}(\phi_n - \mu_k) = 0 \Rightarrow$$

$$\sum_{n=1}^{N} t_{nk}(\phi_n) = \sum_{n=1}^{N} t_{nk}(\mu_k) \Rightarrow \mu_k \underbrace{\sum_{n=1}^{N} t_{nk}}_{N_k} = \sum_{n=1}^{N} t_{nk}\phi_n$$

$$\mu_k = \frac{\sum_{n=1}^{N} t_{nk}\phi_n}{N_k} = \frac{\text{جمع } \phi_n \text{ هایی که عضو کلاس } k \text{ ام اند}}{\text{تعداد داده‌ها از کلاس } k \text{ ام}}$$

$$\boxed{\mu_k = \frac{\sum_{n=1}^{N} \phi_n \,(t_{nk} == 1)}{N_k}}$$

حال برای Σ داریم:

$$L = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left[\ln \pi_k + \ln\left(\frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}}\right) + (-\tfrac{1}{2})(\phi_n-\mu_k)^T \Sigma^{-1} (\phi_n-\mu_k)\right]$$

حال برحسب Σ_k مشتق می‌گیریم.

ادامه 5 ب

استفاده کنید که

$\Sigma^T = \Sigma$, $\frac{\partial}{\partial A} Y^T A Y = Y Y^T$, $\frac{1}{|A|} = |A^{-1}|$

$\frac{\partial}{\partial A} \ln |A| = (A^{-1})^T \rightarrow$ در تمرین ۳ اثبات کردم.

$$\frac{\partial L}{\partial \Sigma_k^{-1}} = \sum_{n=1}^{N} t_{nk} \left[\frac{\partial}{\partial \Sigma^{-1}} \left(\ln |\Sigma^{-1}|^{\frac{1}{2}} \right) + \left(-\tfrac{1}{2}\right) \frac{\partial}{\partial \Sigma^{-1}} (\phi_n - \mu_k)^T \Sigma^{-1} (\phi_n - \mu_k) \right] =$$

$$\frac{\partial L}{\partial \Sigma_k^{-1}} = \sum_{n=1}^{N} t_{nk} \left[\frac{1}{2} \Sigma^T - \frac{1}{2} (\phi_n - \mu_k)^T (\phi_n - \mu_k) \right] = 0$$

$$\Longrightarrow \sum_{n=1}^{N} t_{nk} \left[\cancel{\tfrac{1}{2}} \Sigma - \cancel{\tfrac{1}{2}} (\phi_n - \mu_k)^T (\phi_n - \mu_k) \right] = 0 \Longrightarrow$$

$$\underbrace{\sum_{n=1}^{N} t_{nk} \Sigma}_{N_k \Sigma} = \sum_{n=1}^{N} t_{nk} (\phi_n - \mu_k)^T (\phi_n - \mu_k) \Longrightarrow$$

$$N_k \Sigma = \sum_{n=1}^{N} (\phi_n - \mu_k)^T (\phi_n - \mu_k) \, (t_{nk} == 1) \Longrightarrow$$

$$\boxed{\Sigma_k = \frac{1}{N_k} \times \sum_{n=1}^{N} (\phi_n - \mu_k)^T (\phi_n - \mu_k) (t_{nk} == 1)}$$

Σ_k برای واریانس داده های دسته k ام

μ_k = میانگین داده های دسته k ام.