هوش مصنوعی

نیم‌سال اول ۱۴۰۱-۱۴۰۰

مدرس: دکتر محمدحسین رهبان

دانشکده مهندسی کامپیوتر

مینی‌پروژه چهارم - تئوری

محمدجواد هزاره — شماره دانشجویی: ۹۸۱۰۱۰۷۴

سوال ۱

آ) ویژگی‌ها را شدت نور هر پیکسل در نظر می‌گیریم. بنابراین ورودی عکسی با 3×3 پیکسل خواهد بود که x_{ij} یک است، اگر پیکسل سطر i و ستون j روشن باشد و در غیر این‌صورت صفر خواهد بود. بنابراین با این تعاریف و با استفاده از Naive Bayes، احتمال کلاس‌ها به صورت زیر خواهد بود:

$$\mathbb{P}\{C = A\} = \frac{1}{2}, \qquad \mathbb{P}\{C = B\} = \frac{1}{2}$$

احتمال ویژگی‌ها به شرط کلاس‌ها نیز به صورت زیر خواهد بود که سطر ij هر ماتریس، احتمال $x_{ij} = 1$ به شرط کلاس مورد نظر است:

$$\mathbb{P}\{X = 1|C = A\} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 1 \\ \frac{1}{3} & 0 & 0 \end{pmatrix}, \qquad \mathbb{P}\{X = 1|C = B\} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{2}{3} & 1 & 0 \end{pmatrix}$$

با توجه به صفر شدن احتمال‌ها استفاده از روش Laplace smoothing ر در نظر می‌گیریم. ضریب smoothing را برابر 2 در نظر گرفته و درنتیجه احتمال‌ها به صورت زیر آپدیت خواهند شد:

$$\mathbb{P}\{X = 1|C = A\} = \begin{pmatrix} \frac{3}{7} & \frac{4}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{5}{7} & \frac{5}{7} \\ \frac{3}{7} & \frac{2}{7} & \frac{2}{7} \end{pmatrix}, \qquad \mathbb{P}\{X = 1|C = B\} = \begin{pmatrix} \frac{3}{7} & \frac{3}{7} & \frac{4}{7} \\ \frac{4}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{5}{7} & \frac{2}{7} \end{pmatrix}$$

بنابراین برای داده‌ی ورودی می‌توان احتمال تعلق آن به هر کلاس را محاسبه کرد:

$$\mathbb{P}\{C = A|X_{new}\} = \mathbb{P}\{C = A\}\mathbb{P}\{X_{new}|C = A\} = \frac{1}{2} \times \frac{4}{7} \times \frac{4}{7} \times \frac{4}{7} \times \frac{5}{7} \times \frac{5}{7} \times \frac{5}{7} \times \frac{3}{7} \times \frac{2}{7} \times \frac{5}{7} = \frac{240,000}{2 \times 7^9}$$

$$\mathbb{P}\{C = A|X_{new}\} = \mathbb{P}\{C = A\}\mathbb{P}\{X_{new}|C = A\} = \frac{1}{2} \times \frac{4}{7} \times \frac{3}{7} \times \frac{3}{7} \times \frac{3}{7} \times \frac{3}{7} \times \frac{2}{7} \times \frac{4}{7} \times \frac{5}{7} \times \frac{5}{7} = \frac{64,800}{2 \times 7^9}$$

بنابراین با توجه به این‌که $\mathbb{P}\{C = A|X_{new}\} > \mathbb{P}\{C = B|X_{new}\}$، داده‌ی جدید به کلاس A تعلق خواهد گرفت.

ب) کافیست درختی بسازیم که در هر راس آن براساس روشن یا خاموش بودن پیکسل‌ها بتوان به بچه‌ی راست و چپ آن راس رفت و هر یک از برگ‌ها نیز برچسب یکی از کلاس‌های مسئله را خواهند داشت. با داشتن این درخت، برای یک داده‌ی جدید کافیست براساس روشن و یا خاموش بودن پیکسل‌های آن بر روی درخت حرکت کنیم تا به یک برگ برسیم و برچسب این برگ را به عنوان کلاس این داده خروجی دهیم.

سوال ۲

فرض کنیم داده‌ی ورودی به صورت $\mathcal{D} = \{x_i, y_i\}_{i=1}^{i=m}$ باشد. تعاریف زیر را در نظر می‌گیریم:

$$\begin{cases} B = \min A, \quad A = \{\|w\|_2 \,|\, \forall i \in [m] : y_i \langle w, x_i \rangle \geq 1\} \\ R = \max_i \|x_i\|_2 \end{cases}$$

هم‌چنین فرض می‌کنیم الگوریتم با $w = 0$ شروع می‌کند و در مرحله‌ی آپدیت، بردار وزن به صورت $w^{t+1} = w^t + y_i x_i$ آپدیت می‌شود. (که فرض شده $\eta = 1$؛ برای $\eta \neq 1$ کافیست ورودی‌های مسئله را scale کنیم.) نشان می‌دهیم کران‌بالای دفعاتی که مراحل الگوریتم تکرار خواهند شد برابر با $(RB)^2$ خواهد بود. برای این منظور، بردار $\hat{\boldsymbol{n}}$ را بردار یکه در راستای متناظر با wای در نظر می‌گیریم که B را بدست می‌آورد. به عبارتی خواهیم داشت:

$$\forall i \in \{1, \cdots, m\} : y_i \langle B\hat{\boldsymbol{n}}, x_i \rangle \geq 1 \qquad (*)$$

فرض کنیم الگوریتم k بار اجرا می‌شود. هدف پیدا کردن کرانی برای k است. حال اگر در مرحله‌ی tام، داده‌ی jام اشتباه دسته‌بندی شده باشد، آن‌گاه $\boldsymbol{w^{t+1}} = \boldsymbol{w^t} + y_j \boldsymbol{x_j}$. بنابراین داریم:

$$\begin{aligned} \langle \boldsymbol{w^{t+1}}, B\hat{\boldsymbol{n}} \rangle &= B \langle \boldsymbol{w^t + y_j x_j}, \hat{\boldsymbol{n}} \rangle \\ &= B \left(\langle \boldsymbol{w^t}, \hat{\boldsymbol{n}} \rangle + \langle y_j \boldsymbol{x_j}, \hat{\boldsymbol{n}} \rangle \right) \end{aligned}$$

$$\begin{aligned} \implies \langle \boldsymbol{w^{t+1}}, \hat{\boldsymbol{n}} \rangle &= \langle \boldsymbol{w^t}, \hat{\boldsymbol{n}} \rangle + \frac{1}{B} \langle y_j \boldsymbol{x_j}, B\hat{\boldsymbol{n}} \rangle \\ &\geq \langle \boldsymbol{w^t}, \hat{\boldsymbol{n}} \rangle + \frac{1}{B} \end{aligned}$$

که خط آخر با توجه به $(*)$ نتیجه شده است. حال با توجه به استقرا و این‌که $\boldsymbol{w^1} = \boldsymbol{0}$ ، برای هر t در بازه‌ی ۱ تا k خواهیم داشت: (تمامی نرم‌ها L_2 هستند.)

$$\langle \boldsymbol{w^{t+1}}, \hat{\boldsymbol{n}} \rangle \geq \frac{t}{B} \xRightarrow{\|\hat{\boldsymbol{n}}\|=1} \|\boldsymbol{w^{t+1}}\| \geq \frac{t}{B} \qquad (\star)$$

هم‌چنین برای $\|\boldsymbol{w^{t+1}}\|$ داریم:

$$\begin{aligned} \|\boldsymbol{w^{t+1}}\|^2 &= \|\boldsymbol{w^t} + y_j \boldsymbol{x_j}\|^2 \\ &= \|\boldsymbol{w^t}\|^2 + y_j^2 \|\boldsymbol{x_j}\|^2 + 2 \langle \boldsymbol{w^t}, y_j \boldsymbol{x_j} \rangle \quad (\diamond) \end{aligned}$$

و از آن‌جایی که فرض کرده بودیم داده‌ی jام اشتباه دسته‌بندی شده است، $\langle \boldsymbol{w^t}, y_j \boldsymbol{x_j} \rangle \leq 0$ خواهد بود و هم‌چنین مطابق تعریف الگوریتم برچسب‌ها یک یا منفی یک بودند که نتیجه می‌دهد $y_j^2 = 1$. مطابق تعریف R برای هر i، خواهیم داشت $\|\boldsymbol{x_i}\| \leq R$ ؛ با توجه به این موارد داریم:

$$(\diamond) \implies \|\boldsymbol{w^{t+1}}\|^2 \leq \|\boldsymbol{w^t}\|^2 + R^2$$

با استفاده از استقرا خواهیم داشت:

$$\|\boldsymbol{w^{t+1}}\|^2 \leq t\,R^2 \qquad (\star\star)$$

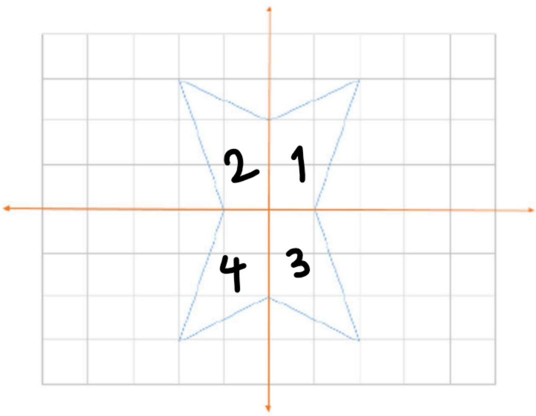
با کنار هم گذاشتن $(\star)$ و $(\star\star)$ و قرار دادن k به جای t خواهیم داشت:

$$\frac{k^2}{B^2} \leq \|\boldsymbol{w^{t+1}}\|^2 \leq k\,R^2 \implies \boxed{k \leq B^2 R^2}$$

بنابراین حداکثر دفعات تکرار مراحل الگوریتم برابر B^2R^2 خواهد بود.

سوال ۳

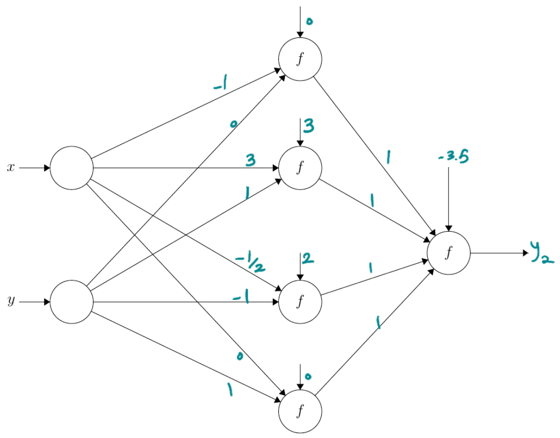
نخست شبکه‌ای که تسخیص دهد ورودی در هر یک از نواحی شماره‌گذاری شده‌ی تصویر زیر باشد را ساخته و سپس با ااجتماع گرفتن بین این شبکه‌ها، خواسته‌ی اصلی مسئله برآورده می‌شود.

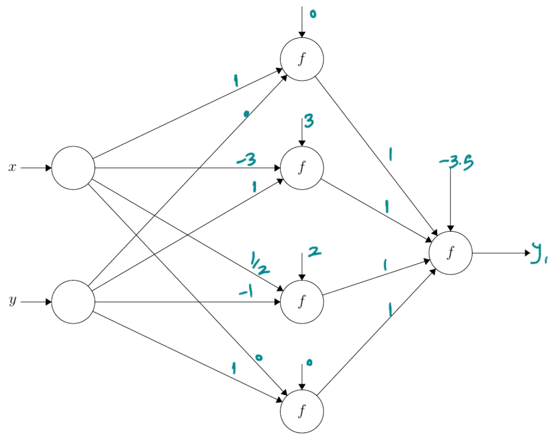


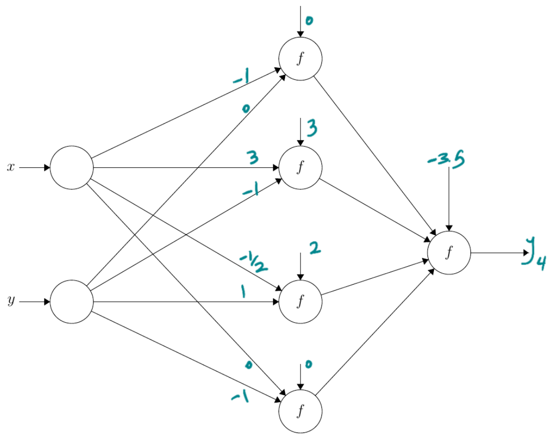
برای هر یک از نواحی نیز، باید تشخیص دهیم ورودی سمت چپ مرزهای آن ناحیه باشد. برای تابع activation نیز از تابع f استفاده می‌کنیم که به صورت زیر تعریف می‌شود:

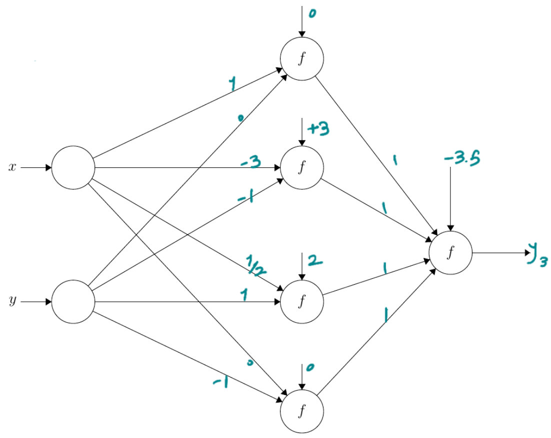
$$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

بنابراین:

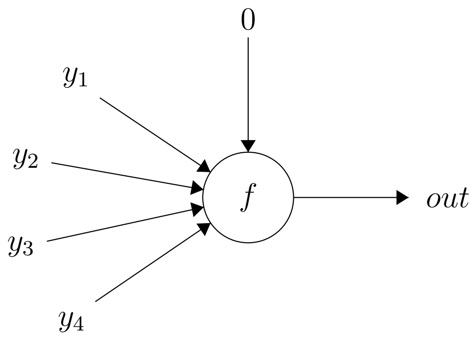








کافیست خروجی‌های y_1 تا y_4 را با یکدیگر or کنیم. بنابراین:



که وزن یال‌های متصل‌کننده‌ی y_iها به این نورون همگی برابر 1 است.