مهسا امینی ۹۸۱۷۸۲۳

تمرین تئوری دوم

**سوال یک**

الف)

$$g(z=\theta^T x) >= 0.5 \rightarrow y=1$$
$$g(z=\theta^T x) < 0.5 \rightarrow y=0$$
$$\theta^T x >= 0 \rightarrow y=1$$
$$\theta^T x < 0 \rightarrow y=0$$

مرز تصمیم برابر است با:

$$\theta^T x = 0$$

محاسبه ی مرز تصمیم:

$$\theta^T x >= 0$$
$$-x_2 + 3 >= 0 \rightarrow y=1$$
$$\theta^T x < 0$$
$$-x_2 + 3 < 0 \rightarrow y=0$$
$$\theta^T x = 0 \rightarrow -x_2 + 3 = 0$$
$$x_2 = 3$$

ب)

$$\theta^T x >= 0$$
$$x_2 + x_1 - 2 >= 0 \rightarrow y=1$$
$$\theta^T x < 0$$
$$x_2 + x_1 - 2 < 0 \rightarrow y=0$$
$$\theta^T x = 0 \rightarrow x_2 + x_1 - 2 = 0$$
$$x_2 = 2 - x_1$$

ج)

$$\sigma(z) = g(z) = \frac{1}{1+e^{-z}}$$
$$h_\theta(x) = g(\theta^T x) = p(y=1 | x)$$
$$p(Y=1|x) + p(Y=0|x) = 1 \rightarrow p(Y=0|x) = 1 - p(Y=1|0)$$
$$g(z=\theta^T x) >= 0.5 \rightarrow y=1$$
$$g(z=\theta^T x) < 0.5 \rightarrow y=0$$

اگر x1=0 و x2=0 باشد در این صورت انتظار داریم y برچسب صفر را بگیرد.

پس داریم:

$$g(z=\theta^T x) < 0.5$$

$$\frac{1}{1+e^{-\theta^T x}} < 0.5$$
$$\frac{1}{1+e^{-(\theta_2 x_2 + \theta_1 x_1 + \theta_0)}} < 0.5$$
$$x_1 = 0, x_2 = 0$$
$$\rightarrow$$
$$\frac{1}{1+e^{-(\theta_0)}} < 0.5$$

پس از حل نامعادله داریم:

$$\theta_0 < 0$$

اگرx1=1 و x2=0 باشد در این صورت انتظار داریم y برچسب یک بگیرد.

پس داریم:

$$\frac{1}{1+e^{-\theta^T x}} >= 0.5$$
$$\frac{1}{1+e^{-(\theta_2 x_2 + \theta_1 x_1 + \theta_0)}} >= 0.5$$
$$x_1 = 1, x_2 = 0$$
$$\rightarrow$$
$$\frac{1}{1+e^{-(\theta_1 + \theta_0)}} >= 0.5$$

پس از حل نا معادله داریم:

$$\theta_1 < -\theta_0$$

اینکار را برای x1=0 و x2=1 تکرار میکنیم و به نتایج مشابه میرسیم:

$$\theta_2 < -\theta_0$$

اگر x1=1 و x2=1 باشد در این صورت نیز y باید برچسب یک بگیرد:

$$\frac{1}{1+e^{-\theta^T x}} >= 0.5$$

$$\frac{1}{1+e^{-(\theta_2 x_2+\theta_1 x_1+\theta_0)}} >= 0.5$$

$$x_1 = 1, x_2 = 1$$

$$\rightarrow$$

$$\frac{1}{1+e^{-(\theta_2+\theta_1+\theta_0)}} >= 0.5$$

پس از حل نا معادله داریم:

$$\theta_0 > -\theta_1 - \theta_2$$

در نهایت بازه های زیر را پیدا کردیم:

$$1)\theta_0 < 0$$

$$2)\theta_1 <= -\theta_0$$

$$3)\theta_2 <= -\theta_0$$

$$4)\theta_0 >= -\theta_1 - \theta_2$$

با توجه به بازه های فوق میتوانیم مقادیر زیر را به پارامتر ها نسبت دهیم:

$$\theta_0 = -1, \theta_1 = 1, \theta_2 = 1$$

چک کردن درستی:

حالت اول:

$$x_1 = 0, x_2 = 0$$

$$\frac{1}{1+e^{-(-1)}} = 0.268$$

$$0.268 < 0.5 \rightarrow y = 0$$

حالت دوم:

$$x_1 = 1, x_2 = 0$$
$$\frac{1}{1+e^{-(1-1)}} = 0.5$$
$$0.5 >= 0.5 \to y = 1$$

حالت سوم:

$$x_1 = 0, x_2 = 1$$
$$\frac{1}{1+e^{-(1-1)}} = 0.5$$
$$0.5 >= 0.5 \to y = 1$$

حالت چهارم:

$$x_1 = 1, x_2 = 1$$
$$\frac{1}{1+e^{-(1+1-1)}} = 0.731$$
$$0.731 >= 0.5 \to y = 1$$

**سوال دو**

قرار است یک بردار بگیریم و احتمال هر کدام را در خروجی داشته باشیم

$$J_{soft\max} = \begin{bmatrix} s_1(1-s_1) & -s_1 s_2 & \dots & -s_1 s_n \\ -s_2 s_1 & s_2(1-s_2) & \dots & -s_2 s_n \\ \dots & \dots & \dots & \dots \\ -s_n s_1 & -s_n s_2 & \dots & s_n(1-s_n) \end{bmatrix}$$

$$s_i = \frac{e^{z_i}}{\sum_{l=1}^{n} e^{zl}}, \forall i = 1, 2, \dots, n$$

میدانیم خروجی تابع softmax احتمال است پس مقادیر آن همگی مثبت است میتوانیم به جای گرفتن مشتق جزئی از خروجی از لگاریتم خروجی مشتق بگیریم:

$$\frac{\partial}{\partial z_j}\log(s_i)=\frac{1}{s_i}.\frac{\partial s_i}{\partial z_j}$$

$$\frac{\partial s_i}{\partial z_j}=s_i.\frac{\partial}{\partial z_j}\log(s_i)$$

$$\log(s_i)=\log(\frac{e^{z_i}}{\sum_{l=1}^{n}e^{zl}})=\log(e^{z_i})-\log(\sum_{l=1}^{n}e^{zl})$$

$$\log(s_i)=z_i-\log(\sum_{l=1}^{n}e^{zl})$$

$$\frac{\partial}{\partial z_j}\log(s_i)=\frac{\partial z_i}{\partial z_j}-\frac{\partial}{\partial z_j}\log(\sum_{l=1}^{n}e^{zl})$$

برای $\frac{\partial z_i}{\partial z_j}$ داریم:

$$f(i=j)$$
$$\frac{\partial z_i}{\partial z_j}=1$$
$$else$$
$$\frac{\partial z_i}{\partial z_j}=0$$

و برای $\frac{\partial}{\partial z_j}\log(\sum_{l=1}^{n}e^{zl})$ داریم:

$$\frac{\partial}{\partial z_j}\log(\sum_{l=1}^{n}e^{zl})=\frac{1}{\sum_{l=1}^{n}e^{zl}}.\left(\frac{\partial}{\partial z_j}\sum_{l=1}^{n}e^{zl}\right)$$

حال از ترکیب هر دو داریم:

$$\frac{\partial}{\partial z_j}\log(s_i)=1\{i=j\}-\frac{1}{\sum_{l=1}^{n}e^{zl}}.\left(\frac{\partial}{\partial z_j}\sum_{l=1}^{n}e^{zl}\right)$$

$$\frac{\partial}{\partial z_j}\sum_{l=1}^{n} e^{zl} = \frac{\partial}{\partial z_j}[e^{z_1} + e^{z_2} + ... + e^{z_j} + .... + e^{z_n}] = \frac{\partial}{\partial z_j}[e^{z_j}] = e^{z_j}$$

$$\frac{\partial}{\partial z_j}\log(s_i) = 1\{i = j\} - \frac{e^{z_j}}{\sum_{l=1}^{n} e^{zl}} = 1\{i = j\} - s_j$$

پس بر طبق این رابطه $\frac{\partial s_i}{\partial z_j} = s_i . \frac{\partial}{\partial z_j}\log(s_i)$ که قبلا به آن رسیدم داریم:

$$\frac{\partial s_i}{\partial z_j} = s_i .(1\{i = j\} - s_j)$$

حال ماتریسی که در ابتدا تعریف کردیم را بازنویسی میکنیم:

$$J_{soft\max} = \begin{bmatrix} s_1(1-s_1) & -s_1s_2 & ... & -s_1s_n \\ -s_2s_1 & s_2(1-s_2) & ... & -s_2s_n \\ ... & ... & ... & ... \\ -s_ns_1 & -s_ns_2 & ... & s_n(1-s_n) \end{bmatrix}$$

**سوال سه**

قسمت الف)

$$a_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, y_1 = -1$$

$$a_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, y_2 = -1$$

$$a_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, y_3 = 1$$

$$L(w,b,\alpha)=\frac{1}{2}\|w\|_2^2-\sum_{i=1}^{n}\alpha_i(y^i(x^iw+b)-1)=$$

$$\frac{1}{2}w^Tw-\sum_{i=1}^{n}\alpha_i y^i x^i w-b\sum_{i=1}^{n}\alpha_i y_i+\sum_{i=1}^{n}\alpha_i$$

$$1)\ \frac{dL}{dw}=w-\sum_{i=1}^{n}\alpha_i y^i x^i \rightarrow w=\sum_{i=1}^{n}\alpha_i y^i x^i$$

$$2)\ \frac{dL}{db}=\sum_{i=1}^{n}\alpha_i y^i$$

از رابطه یک داریم:

$$w^*=-\alpha_1\begin{bmatrix}1\\4\end{bmatrix}-\alpha_2\begin{bmatrix}2\\3\end{bmatrix}+\alpha_3\begin{bmatrix}4\\5\end{bmatrix}$$

از رابطه ی دو داریم:

$$-\alpha_1-\alpha_2+\alpha_3=0$$

$$\max_{\alpha_1,\alpha_2,\alpha_3} L(w,b,\alpha)=\sum\alpha_i-\frac{1}{2}\sum y^i y^j\alpha_i\alpha_j(x^i)^T x^j-b\sum\alpha_i y^i$$

$$x_1^Tx_1=17, x_1^Tx_2=14, x_1^Tx_3=24$$
$$x_2^Tx_1=14, x_2^Tx_2=13, x_2^Tx_3=23$$
$$x_3^Tx_1=24, x_3^Tx_2=23, x_3^Tx_3=41$$

$$\max_{\alpha_1,\alpha_2,\alpha_3}\rightarrow\alpha_1+\alpha_2+\alpha_3-\frac{1}{2}[17\alpha_1^{\ 2}+28\alpha_1\alpha_2-48\alpha_1\alpha_3+13\alpha_2^{\ 2}-46\alpha_2\alpha_3+41\alpha_3^{\ 2}]$$

از قبل داشتیم:

$$-\alpha_1-\alpha_2+\alpha_3=0$$

پس داریم:

$$\alpha_1+\alpha_2=\alpha_3$$

پس از جایگذاری به عبارت زیر میرسیم:

$$\max_{\alpha_1,\alpha_2}\rightarrow-5\alpha_1^{\ 2}-8\alpha_1\alpha_2-4\alpha_2^{\ 2}+2\alpha_1+2\alpha_2$$

$$\frac{dL}{d\alpha_1} = -10\alpha_1 - 8\alpha_2 + 2 = 0$$
$$\frac{dL}{d\alpha_2} = -8\alpha_1 - 8\alpha_2 + 2 = 0$$

پس از حل دستگاه:

$$\alpha_1 = 0, \alpha_2 = 0.25, a_3 = 0.25$$

$$w^* = 0\begin{bmatrix}1\\4\end{bmatrix} - 0.25\begin{bmatrix}2\\3\end{bmatrix} + 0.25\begin{bmatrix}4\\5\end{bmatrix} = \begin{bmatrix}0.5\\0.5\end{bmatrix}$$

$$y_2(x_2^T w + b^*) = 1$$
$$b^* = -1 - \begin{bmatrix}2 & 3\end{bmatrix}\begin{bmatrix}0.5\\0.5\end{bmatrix} = -3.5$$

$$m = \frac{2}{\|w\|} = 2\sqrt{2}$$

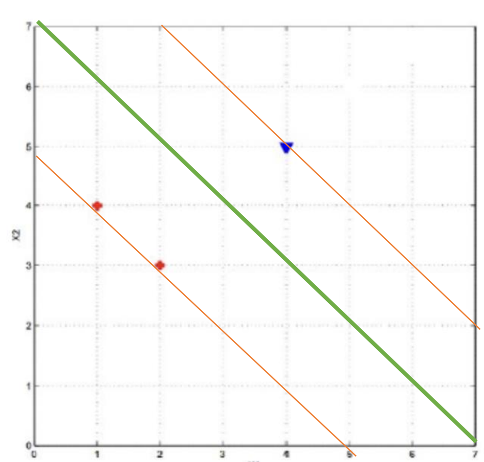
ب)

معادله ی تصمیم: خط سبز خط تصمیم و نارنجی مشخص کننده ی نقاط پشتیبان است که به فاصله $\sqrt{2}$ از مرز تصمیم قرار گرفته.

$$x^T w^* + b^* = \begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}0.5\\0.5\end{bmatrix} - 3.5 = 0.5x_1 + 0.5x_2 = 3.5$$

$$0.5x_2 = -0.5x_1 + 3.5$$
$$x_2 = -x_1 + 7$$

$$m = 2\sqrt{2}$$
$$\frac{m}{2} = \sqrt{2} = 1.4$$



**سوال چهار**

الف)

$$G(D) = 1 - \sum_{i=1}^{k} {p_i}^2$$

$$1 - ((\frac{10}{20})^2 + (\frac{10}{20})^2) = \frac{1}{2}$$

ب)

$$binary \quad partition$$

$$G_A(D) = \frac{|D_1|}{D} G(D_1) + \frac{|D_2|}{D} G(D_2)$$

$$G(D_1) = 1 - (1^2 + 0^2) = 0$$

$$G(D_2) = 1 - ((\frac{10}{19})^2 + (\frac{9}{19})^2) = 0.4986$$

$$G_{customerID}(D) = \frac{1}{20}.(0) + \frac{19}{20}(0.4986) = 0.47367$$

به همین ترتیب برای تک تک id ها میتوانیم محاسبه کنیم که به نتایج مشابه خواهیم رسید.

$$\gamma(customerID, D) = G(D) - G_{customerID}(D)$$
$$\gamma(customerID, D) = 0.5 - 0.47367 = 0.02633$$

ج)

$$G_A(D) = \frac{|D_1|}{D} G(D_1) + \frac{|D_2|}{D} G(D_2)$$

$$G(D_1) = 1 - ((\frac{6}{10})^2 + (\frac{4}{10})^2) = 0.48$$

$$G(D_2) = 1 - ((\frac{6}{10})^2 + (\frac{4}{10})^2) = 0.48$$

$$G_{gender}(D) = \frac{10}{20}.(0.48) + \frac{10}{20}(0.48) = 0.48$$
$$\gamma(gender, D) = G(D) - G_{gender}(D)$$
$$\gamma(gender, D) = 0.5 - 0.48 = 0.02$$

د)

$$multiway \quad split$$

$$G_A(D)=\frac{|D_1|}{D}G(D_1)+\frac{|D_2|}{D}G(D_2)+\frac{|D_3|}{D}G(D_3)$$

$$G(D_1)=1-((\frac{3}{4})^2+(\frac{1}{4})^2)=0.375$$

$$G(D_2)=1-((\frac{8}{8})^2+(\frac{0}{8})^2)=0$$

$$G(D_3)=1-((\frac{1}{8})^2+(\frac{7}{8})^2)=0.21875$$

$$G_{carType}(D)=\frac{4}{20}.(0.375)+\frac{8}{20}(0)+\frac{8}{20}(0.21875)=0.1625$$

$$\gamma(carType,D)=G(D)-G_{carTyoe}(D)$$

$$\gamma(carType,D)=0.5-0.1625=0.3375$$

و)

$$multiway \quad split$$

$$G_A(D)=\frac{|D_1|}{D}G(D_1)+\frac{|D_2|}{D}G(D_2)+\frac{|D_3|}{D}G(D_3)+\frac{|D_4|}{D}G(D_4)$$

$$G(D_1)=1-((\frac{2}{5})^2+(\frac{3}{5})^2)=0.48$$

$$G(D_2)=1-((\frac{3}{7})^2+(\frac{4}{7})^2)=0.4898$$

$$G(D_3)=1-((\frac{2}{4})^2+(\frac{2}{4})^2)=0.5$$

$$G(D_4)=1-((\frac{2}{4})^2+(\frac{2}{4})^2)=0.5$$

$$G_{carType}(D)=\frac{5}{20}.(0.48)+\frac{7}{20}(0.4898)+\frac{4}{20}(0.5)+\frac{4}{20}(0.5)=0.46743$$

$$\gamma(shirtSize,D)=G(D)-G_{shirtSize}(D)$$

$$\gamma(shirtSize,D)=0.5-0.46743=0.03257$$

ه) بین این سه ویژگی carType را انتخاب میکنیم چرا که $\gamma(A,D)$ را ماکسیمایز میکند و شاخص جینی آن از بقیه کمتر است.

ی) این ویژگی قدرت پیش بینی ندارد زیرا مشتریان جدید به آی دی ها ی مشتریان جدید اختصاص داده می شوند.

**سوال پنج**

الف)

$$E(D) = -\sum_{i=1}^{k} p_i \log_1(p_i)$$

$$-(\frac{4}{9}\log_1(\frac{4}{9}) + \frac{5}{9}\log_2(\frac{5}{9}))$$
$$-(-0.52 - 0.471) = 0.9911$$

ب)

$$E_A(D) = \sum_{j=1}^{m} \frac{|D_j|}{D}.E(D_j)$$
$$\alpha(A,D) = E(D) - E_A(D)$$
$$E(D_1) = -(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}) = 0.81123$$
$$E(D_2) = -(\frac{4}{5}\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5}) = 0.72192$$
$$E_{\alpha_1}(D) = \frac{4}{9}0.81123 + \frac{5}{9}0.72192 = 0.7616$$
$$\alpha(\alpha_1, D) = 0.9911 - 0.7616 = 0.2295$$

$$E(D_1) = -(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}) = 1$$
$$E(D_2) = -(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}) = 0.971$$
$$E_{\alpha_2}(D) = \frac{4}{9} + \frac{5}{9} * 0.971 = 0.9838$$
$$\alpha(\alpha_2, D) = 0.9911 - 0.9838 = 0.0072$$

ج)

1- برای نقطه یک داریم (در این حالت با دو مقایسه میکنیم):

$$t_1 = \frac{1}{9}\log_2(1) = 0$$
$$t_2 = \frac{8}{9}(-\frac{3}{8}\log(\frac{3}{8}) - \frac{5}{8}\log(\frac{5}{8})) = 0.8484$$
$$t_1 + t_2 = 0.8484$$

$$\alpha(\alpha_3, D) = 0.9911 - 0.8484 = 0.1427$$

2- برای نقطه سه داریم ( دراین حالت با سه و نیم مقایسه میکنیم):

$$t_1 = \frac{1}{9}(-\log_2(\frac{1}{2}) - \log_2(\frac{1}{2})) = 0.2222$$
$$t_2 = \frac{7}{9}(-\frac{4}{7}\log(\frac{4}{7}) - \frac{3}{7}\log(\frac{3}{7})) = 0.76631$$
$$t_1 + t_2 = 0.9885$$

$$\alpha(\alpha_3, D) = 0.9911 - 0.9885 = 0.0026$$

3- برای نقطه ی چهار داریم( دراین حالت با چهار و نیم مقایسه میکنیم):

$$t_1 = \frac{3}{9}(-\frac{2}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{1}{3})) = 0.3061$$
$$t_2 = \frac{6}{9}(-\frac{4}{6}\log(\frac{4}{6}) - \frac{2}{6}\log(\frac{2}{6})) = 0.6122$$
$$t_1 + t_2 = 0.9183$$
$$\alpha(\alpha_3, D) = 0.9911 - 0.9183 = 0.0728$$

4- برای نقطه ی پنج داریم( دراین حالت با پنج و نیم مقایسه میکنیم):

$$t_1 = \frac{5}{9}(-\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5})) = 0.5394$$
$$t_2 = \frac{4}{9}(-\frac{2}{4}\log(\frac{2}{4}) - \frac{2}{4}\log(\frac{2}{4})) = 0.4444$$
$$t_1 + t_2 = 0.9838$$
$$\alpha(\alpha_3, D) = 0.9911 - 0.9183 = 0.0072$$

5- برای نقطه ی شش داریم( دراین حالت با شش و نیم مقایسه میکنیم):

$$t_1 = \frac{6}{9}(-\frac{3}{6}\log_2(\frac{3}{6}) - \frac{3}{6}\log_2(\frac{3}{6})) = 0.6666$$
$$t_2 = \frac{3}{9}(-\frac{1}{3}\log(\frac{1}{3}) - \frac{2}{3}\log(\frac{2}{3})) = 0.3061$$
$$t_1 + t_2 = 0.9727$$
$$\alpha(\alpha_3, D) = 0.9911 - 0.9183 = 0.0183$$

6- برای نقطه ی هفت داریم( در این حالت با هفت و نیم مقایسه میکنیم):

$$t_1 = \frac{8}{9}(-\frac{4}{8}\log_2(\frac{4}{8}) - \frac{4}{8}\log_2(\frac{4}{8})) = 0.8888$$
$$t_2 = \frac{1}{9}(-\log_2(1)) = 0$$
$$t_1 + t_2 = 0.8888$$
$$\alpha(\alpha_3, D) = 0.9911 - 0.9183 = 0.1022$$

زمانی که نقاطی که کمتر از دو هستند را در یک دسته گذاشتیم و بقیه را در دسته ی دیگر به بیشترین $\alpha(\alpha_3, D)$ رسیدیم.

د) بین یک و سه ، یک بهتر است چرا که $\alpha$ بیشتری دارد و بین سه و دو ، سه بهتر است زیرا $\alpha$ بیشتری دارد. و در یک یک از همه بهتر است.

و) یک بهتر است خطای آن 2/9 است و از دومی کمتر است.

ه)

$$G_A(D) = \frac{|D_1|}{D}G(D_1) + \frac{|D_2|}{D}G(D_2)$$
$$G(D_1) = 1 - ((\frac{1}{4})^2 + (\frac{3}{4})^2) = 0.375$$
$$G(D_2) = 1 - ((\frac{1}{5})^2 + (\frac{4}{5})^2) = 0.32$$
$$G_{\alpha_1}(D) = \frac{4}{9}.(0.375) + \frac{5}{9}(0.32) = 0.3444$$

$$G_A(D) = \frac{|D_1|}{D}G(D_1) + \frac{|D_2|}{D}G(D_2)$$
$$G(D_1) = 1 - ((\frac{2}{5})^2 + (\frac{3}{5})^2) = 0.48$$
$$G(D_2) = 1 - ((\frac{2}{4})^2 + (\frac{2}{4})^2) = 0.5$$
$$G_{\alpha_2}(D) = \frac{5}{9}.(0.48) + \frac{4}{9}(0.5) = 0.4888$$

شاخص یک کمتر است پس مناسب تر است.

سوال شش

$$Y=\{y_1,y_2,...,y_c\}\to c \quad classes$$
$$X=\{x_1,x_2,...,x_k\}\to k \quad attribute$$

قبل از تقسیم به گره های جانشین داریم:

$$E(Y)=-\sum_{j=1}^{c}P(y_i)\log_2 P(y_i)$$

بر اساس قانون احتمال کل میدانیم:

$$P(y_i)=\sum_{i=1}^{k}P(x_i,y_i)$$

حال بر طبق این قانون:

$$E(Y)=-\sum_{j=1}^{c}P(y_i)\log_2 P(y_i)=\sum_{j=1}^{c}\sum_{i=1}^{k}P(x_i,y_i)\log_2 P(y_i)$$

بعد از تقسیم کردن آنتروپی برای هر child :

$$E(Y\mid x_i)=-\sum_{j=1}^{c}P(y_i\mid x_i)\log_2 P(y_i\mid x_i)$$

همچنین داریم:

$$E(Y\mid X)=-\sum_{i=1}^{k}P(x_i)E(Y\mid x_i)$$

با جایگذاری $E(Y\mid x_i)$ در رابطه بالا:

$$E(Y\mid X)=-\sum_{i=1}^{k}\sum_{j=1}^{c}\ P(x_i)P(y_i\mid x_i)\log_2 P(y_i\mid x_i)$$

میدانیم:

$$P(x_i,y_i)=P(y_i\mid x_i)\times P(x_i)$$

پس $E(Y\mid X)$ را دوباره بازنویسی میکنیم:

$$E(Y\mid X)=-\sum_{i=1}^{k}\sum_{j=1}^{c}P(x_i,y_i)\log_2 P(y_i\mid x_i)$$

برای اثبات خواسته ی مسئله باید این را ثابت کنیم که:

$$E(Y|X) \le E(Y)$$

$$E(Y|X) - E(Y) =$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{c} P(x_i, y_i)\log_2 P(y_i | x_i) + \sum_{i=1}^{k}\sum_{j=1}^{c} P(x_i, y_i)\log_2 P(y_i)$$

$$\sum_{i=1}^{k}\sum_{j=1}^{c} P(x_i, y_i)\log_2 \frac{P(y_i)}{P(y_i | x_i)} =$$

$$\sum_{i=1}^{k}\sum_{j=1}^{c} P(x_i, y_i)\log_2 \frac{P(x_i)P(y_i)}{P(x_i, y_i)}$$

با این شرایط میدانیم:

$$\sum_{k=1}^{n} a_k \log(t_k) \le \log(\sum_{k=1}^{n} a_k t_k)$$

$$E(Y|X) - E(Y) \le \log_2(\sum_{i=1}^{k}\sum_{j=1}^{c} P(x_i, y_i)\frac{P(x_i)P(y_i)}{P(x_i, y_i)})$$

$$= \log(\sum_{i=1}^{k} p(x_i)\sum_{j=1}^{c} p(y_i)) = 0$$

پس:

$$E(Y|X) \le E(Y)$$

**سوال هفت**

در soft margin svm ما میخواهیم مقدار عبارت زیر را مینموم کنیم

$$L = \frac{1}{2}\|w\|^2 + C(\# mistakes)$$

که c یک تریدآف برای ماکسیمم کردن مارجین و مینیموم کردن خطا است. از آنجایی که خطاهای ما یکسان نیستند و داده ها در فاصله ی متفاوتی از مرز تصمیم قرار گرفته اند پس باید یک پنالتی به آن اضافه کرد.

پس برای هر داده ی x_i یک $\varepsilon\_i$ در نظر میگیریم که اگر اشتباه دسته بندی شده باشد $\varepsilon\_i$ برابر فاصله ی آن داده از حاشیه ی کلاس مربوطه است. و اگر درست دسته بندی شده باشد برابر صفر است. و هر چه داده ای که اشتباه دسته بندی شده فاصله ی بیشتری از حاشیه داشته باشد باید پنالتی بیشتری بگیرد پس با توجه به موارد گفته شده x_i باید شرط زیر را ارضا کند:

$$y_i(w.x_i + b) \ge 1 - \varepsilon_i$$

اگر $\varepsilon\_i$ بین صفر و یک باشد یعنی درست دسته بندی میشود ولی داخل مارجین است اگر بزرگ تر از یک باشد یعنی دسته بندی درست انجام نمیشود هرچقدر مقدار $\varepsilon\_i$ بیشتر باشد جریمه بیشتر میشود.

با توجه به محدودیت های گفته شده هدف ما مینیموم کردن تابع زیر است:

$$L = \frac{1}{2} \| w \|^2 + C \sum_i \varepsilon_i + \sum_i \lambda_i (y_i (w.x_i + b) - 1 + \varepsilon_i)$$

که برای رسیدن به این از ضرایب لاگرانژ استفاده کرده ایم.