دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

درس مبانی یادگیری ماشین

تکلیف تئوری اول

مهلت تحویل: ۲۰ آبان ۱۴۰۲

سوال ۱

الف) غلط- استفاده از روش‌های منتظم‌سازی ارتباطی با بهینه‌سازی مدل و کمک به فرار از مینیمم محلی ندارد و برای کنترل مقادیر وزن‌ها و جلوگیری از بیش‌برازش است.

ب) صحیح- می‌دانیم که در بهینه‌سازی خطای جمع مربعات با درنظر گرفتن بردار ویژگی‌های ϕ بردار وزن‌های بیشینه درستنمایی به صورت زیر حاصل می‌شود

$$\begin{aligned} E_D(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\left\{t_n - \mathbf{w}^{\mathrm{T}}\phi\left(\mathbf{x}_n\right)\right\}^2 \\ \nabla E_D(\mathbf{w}) &= \sum_{n=1}^{N} -\left\{t_n - \mathbf{w}^{\mathrm{T}}\phi\left(\mathrm{x}_n\right)\right\}\phi\left(\mathrm{x}_n\right) \\ 0 &= \sum_{n=1}^{N} t_n\phi\left(\mathrm{x}_n\right) - \mathbf{w}\left(\sum_{n=1}^{N}\phi\left(\mathrm{x}_n\right)^{\mathrm{T}}\phi\left(\mathrm{x}_n\right)\right) \\ 0 &= \Phi^T\mathbf{t} - \Phi^T\Phi\mathbf{w} \end{aligned} \tag{۱}$$

که ماتریس طراحی با در نظر گرفتن M ویژگی و N داده به صورت زیر است

$$\Phi = \begin{pmatrix} \phi_0\left(\mathrm{x}_1\right) & \phi_1\left(\mathrm{x}_1\right) & \cdots & \phi_{M-1}\left(\mathrm{x}_1\right) \\ \phi_0\left(\mathrm{x}_2\right) & \phi_1\left(\mathrm{x}_2\right) & \cdots & \phi_{M-1}\left(\mathrm{x}_2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0\left(\mathrm{x}_N\right) & \phi_1\left(\mathrm{x}_N\right) & \cdots & \phi_{M-1}\left(\mathrm{x}_N\right) \end{pmatrix} \tag{۲}$$

بردار وزن‌های به دست آمده با بیشینه درستنمایی به صورت زیر است

$$w_{MLE} = (\Phi^T\Phi)^{-1}\Phi^T t \tag{۳}$$

که از آن رابطه‌ی زیر حاصل می‌شود

$$\mathbf{0} = \Phi^T(t - \Phi w_{MLE}), \quad y = \Phi w_{MLE} \tag{۴}$$

که بدیهی است برای صفر شدن (کمینه شدن) خطای جمع مربعات، لازم است اختلاف بردار t و y بر صفحه‌ی ϕ_i ها عمود باشد یا به عبارت دیگر حاصل ضرب برداری آن‌ها صفر شود.

ج) صحیح- هرچه مدل ساده‌تر باشد، باعث می‌شود که مدل روی داده‌ها به خوبی آموزش نبیند که در این صورت بایاس مدل زیاد شده و مشکل زیربرازش رخ می‌دهند. در این حالت واریانس مدل کم خواهد بود.

د) غلط- هدف از مجموعه داده ارزیابی، ابرتنظیم پارامترهای مدل یادگیری ماشین برای عملکرد بهتر روی داده‌های دیده نشده‌ی مجموعه‌ی آزمایش است. اگر داده های ارزیابی از توزیع متفاوتی بدست آید، ممکن است تخمین دقیقی از عملکرد مدل در دنیای واقعی ارائه نکند که این می تواند منجر به ارزیابی بیش از حد خوش بینانه یا بدبینانه از قابلیت های مدل شود.

هـ) غلط- اگر مجموعه داده آموزشی موجود متعادل نباشد، نمونه برداری تصادفی انتخاب خوبی نیست. باید سعی کنیم که طوری نمونه برداری کنیم که مجموعه ها متعادل باشند و یا اول مجموعه داده را با برخی روش های نمونه افزایی[1] بزرگ کنیم، طوری

[1] Upsampling

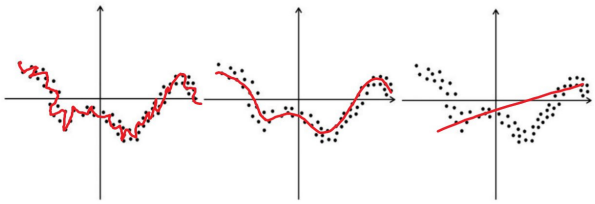
که کلاس های موجود متعادل شوند و سپس به صورت تصادفی نمونه برداری کنیم. در حالت دیگر، اگر ورودی مدل از جنس سری زمانی باشد منطقی نیست که به صورت تصادفی از داده های موجود نمونه برداری کنیم بلکه باید از ابتدای زمان موجود تا زمان t_0 را برای مجموعه داده آموزشی انتخاب کنیم و از زمان t_0 به بعد را به همین ترتیب بین مجموعه ارزیابی و آزمایش تقسیم کنیم. علت این امر این است که در صورت نمونه برداری تصادفی، قدرت برون یابی مدل به خوبی سنجیده نمی شود.

و) غلط- به این سبب که در بخش زیادی از داده ها $\hat{y} = y$ است لذا استفاده از تابع هزینه MAE به افزایش تعداد نقاط مشتق ناپذیر گشته و به دلیل لزوم استفاده از $subgradient$ بهتر است از MSE عنوان تابع هزینه استفاده نماییم تا با نقاط مشتق ناپذیر مواجه نشویم

ز) غلط- مدل برروی همه‌ی داده‌ها آموزش داده شده، سپس نتایج آن به عنوان گزارش نهایی ارائه می‌شود.

سؤال ۲

در شکل زیر داریم



سؤال ۳

ستون‌های a و b و c به ترتیب برای مقادیر $\infty-$ و ۱۸- و صفر هستند. زیرا در حالتی که $\ln \lambda = -\infty$ مقدار λ نزدیک صفر بوده و مقادیر وزن‌ها کنترل نمی‌شوند. بنابراین مدل دچار مشکل بیش‌برازش می‌شود. همچنین در حالتی که $\ln \lambda = 0$ است، مقدار λ برابر ۱ بوده در نتیجه برای بزرگ شدن مقدار وزن‌ها پنالتی بزرگتری در نظر می‌گیرد و وزن‌ها نزیک به صفر خواهند بود. همچنین در حالتی که $\ln \lambda = -18$ ، مقدار λ بین صفر و ۱ خواهد بود که سبب می‌شود پنالتی منطقی‌تری در زیاد شدن قدر مطلق وزن‌ها در نظر بگیرد.

سؤال ۴

تابع چگالی توزیع احتمال یکنواخت با پارامتر θ به صورت زیر است

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq 1 \\ 0, & \text{در غیر اینصورت} \end{cases}$$

با توجه به n داده‌ی موجود تابع درستنمایی به صورت زیر است

$$l(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 \leq x \leq 1 \\ 0, & \text{در غیر اینصورت} \end{cases}$$

از آنجا که تابع درستنمایی نزولی است، برای بیشینه شدن آن لازم است که مقدار θ کمینه باشد. اما با توجه به داشتن نمونه‌ها، مقدار کمینه θ نمی‌تواند به گونه‌ای باشد که همه‌ی نمونه‌ها را شامل نشود. بنابراین کمینه مقدار θ که نمونه‌ها را شامل شود به صورت زیر است

$$\theta_{MLE} = \max(x_1, ..., x_n)$$

سؤال ۵

$$Posterior \propto Likelihood \times Prior$$

محاسبه لگاریتم تابع درستنمایی به صورت زیر است

$$\begin{aligned} L(p) &= \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i} \\ logL(p) &= \sum_{i=1}^{N} x_i \log(p) + (1-x_i)\log(1-p) \\ &= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i) \end{aligned}$$

محاسبه لگاریتم تابع پیشین

$$\log f(p|\alpha,\beta) = \log\left(\frac{1}{B(\alpha,\beta)}\right) + (\alpha-1)\log(x) + (\beta-1)\log(1-x)$$

محاسبه MAP به صورت زیر است

$$\log(Posterior) \propto \log(Likelihood) + \log(Prior)$$

$$\begin{aligned} l(p) + \log(f(p|\alpha,\beta)) &= \log(p)\sum_i x_i + \log(1-p)\sum_i(1-x_i) - \log\left(B(\alpha,\beta)\right) + (\alpha-1)\log(x) + (\beta-1)\log(1-x) \\ &= \log(p)\left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) + \log(1-p)\left(\beta - 1 + n - \sum_{i=1}^{n} xi\right) - \log\left(B(\alpha,\beta)\right) \end{aligned}$$

حال از عبارت بالا نسبت به p مشتق گرفته برابر صفر قرار می‌دهیم

$$\begin{aligned} &\frac{\partial[l(p) + \log(f(p \mid \alpha,\beta))]}{\partial p} = 0 \\ &\rightarrow \frac{(\alpha - 1 + (\sum_{i=1}^{n} xi))}{p} - \frac{(\beta - 1 + n - \sum_{i=1}^{n} xi)}{1-p} = 0 \\ &\rightarrow \frac{(1-p)\,(\alpha - 1 + (\sum_{i=1}^{n} xi)) - p\,(\beta - 1 + n - \sum_{i=1}^{n} xi)}{p(1-p)} = 0 \\ &\rightarrow \left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) - p\left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) - p\left(\beta - 1 + n - \sum_{i=1}^{n} xi\right) = 0 \\ &\rightarrow \left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) - p\left(\left(\alpha - 1 + \left(\sum_{i=1}^{n} xi\right)\right) + \left(\beta - 1 + n - \sum_{i=1}^{n} xi\right)\right) = 0 \\ &\rightarrow p = \frac{(\alpha - 1 + (\sum_{i=1}^{n} xi))}{(\alpha - 1 + (\sum_{i=1}^{n} xi)) + (\beta - 1 + n - \sum_{i=1}^{n} xi)} = \frac{(\alpha - 1 + (\sum_{i=1}^{n} xi))}{\alpha + \beta + n - 2} \end{aligned}$$

سؤال ۶

$$P(x_1, ..., x_{10}|\theta) = \prod_{i=1}^{10} P(x_i|\theta)$$
$$= \frac{\theta}{2} \times \frac{(1-\theta)}{2} \times \frac{\theta}{4} \times \frac{3(1-\theta)}{4} \times \frac{\theta}{4} \times \frac{\theta}{2} \times \frac{3(1-\theta)}{4} \times \frac{\theta}{4} \times \frac{(1-\theta)}{2} \times \frac{3(1-\theta)}{4}$$

$$\log(P(x_1, ..., x_{10}|\theta)) = \frac{\theta^5(1-\theta^5)27}{2^{16}}$$
$$= 5\log\theta + 5\log(1-\theta) + 3\log 3 - 16\log 2$$

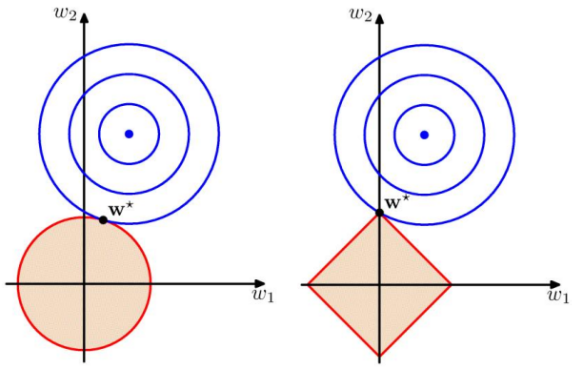
حال نسبت به θ مشتق گرفته برابر صفر قرار می‌دهیم

$$\frac{dP(x_1, ..., x_{10}|\theta)}{d\theta} = \frac{5}{\theta} - \frac{5}{1-\theta} = 0$$

$$\theta_{MLE} = 0.5$$

سؤال ۷

ستون A متعلق به تابع هزینه‌ی $Ridge$ و ستون B متعلق به تابع هزینه‌ی $Lasso$ است. همانطور که در شکل زیر نشان داده شده، احتمال برخورد کانتورها با ترم منتظم سازی $Lasso$ بر روی محورهای وزن‌ها بیشتر است که منجر می‌شود بعضی دیگر وزن‌ها مقدار صفر پیدا کنند. مزیت این تابع هزینه در کم شدن تعداد وزن‌ها و کاهش حجم محاسبات است. همچنین احتمال برخورد کانتورها با ترم منتظم سازی درجه دو در تابع هزینه‌ی $Ridge$ در نقاط روی محور وزن‌ها کمتر است و وزن‌های خیلی کمتری مقدار صفر پیدا می‌کنند. مزیت این تابع هزینه در مشتق پذیری آن است.



سؤال ۸

الف) محاسبه‌ی خطاها به صورت زیر است

$$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -59.5 + -0.15 * 41 + 0.60 * 138 = 17.15$$

$$\hat{y}^{(2)} = w[0] + w[1] * 42 + w[2] * 153 = -59.5 + -0.15 * 42 + 0.60 * 153 = 26.00$$

$$\hat{y}^{(3)} = w[0] + w[1] * 37 + w[2] * 151 = -59.5 + -0.15 * 37 + 0.60 * 151 = 25.55$$

$$\hat{y}^{(4)} = w[0] + w[1] * 46 + w[2] * 133 = -59.5 + -0.15 * 46 + 0.60 * 133 = 13.40$$

$$MSE = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{8} \sum_{i=1}^{4} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 =$$

$$\frac{1}{8} \sum_{i=1}^{4} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{8} \left[(17.15 - 37.99)^2 + (26.00 - 47.34)^2 + (25.55 - 44.38)^2 \right. \quad (۵)$$

$$\left. + (13.40 - 28.17)^2 \right] = 182.8028$$

$$MAE = \frac{1}{m} \sum_{i=1}^{m} \left| \hat{y}^{(i)} - y^{(i)} \right| = \frac{1}{8} \sum_{i=1}^{4} \left| \hat{y}^{(i)} - y^{(i)} \right| =$$

$$\frac{1}{4} \sum_{i=1}^{4} \left| \hat{y}^{(i)} - y^{(i)} \right| = \frac{1}{4} [|17.15 - 37.99| + |26.00 - 47.34| + |25.55 - 44.38| + |13.40 - 28.17|]$$

$$= 18.945$$

ب) مراحل به روزرسانی به شرح زیر است

$$L(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$L_i(w) = \frac{1}{2}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial L_i(w)}{\partial w_j} = \left(\hat{y}^{(i)} - y^{(i)}\right)x_j^{(i)}, \quad i \in \{1,2\}$$

$$\frac{\partial L_i(w)}{\partial w_0} = \left(\hat{y}^{(i)} - y^{(i)}\right)$$

$$w_0, w_1, w_2 = [-59.5, -0.15, 0.6]$$

$$\frac{\partial L_i(w)}{\partial w_1} = (25.55 - 44.38) * 37 = -696.71$$

$$\frac{\partial L_i(w)}{\partial w_2} = (25.55 - 44.38) * 151 = -2843.33$$

$$\frac{\partial L_i(w)}{\partial w_0} = (25.55 - 44.38) = -18.83$$

$$w[0] = -59.50 - 0.1 * (-18.83) = -57.617$$

$$w[1] = -0.15 - 0.1 * (-696.71) = 69.521$$

$$w[2] = 0.60 - 0.1 * (-2843.33) = 284.933$$

(۶)

$$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -57.617 + 69.521 * 41 + 284.933 * 138$$

$$= 42113.498$$

$$\frac{\partial L_i(w)}{\partial w_1} = (42113.498 - 37.99) * 41 = 1725095.828$$

$$\frac{\partial L_i(w)}{\partial w_2} = (42113.498 - 37.99) * 138 = 5806420.104$$

$$\frac{\partial L_i(w)}{\partial w_0} = (42113.498 - 37.99) = 42075.508$$

$$w[0] = -57.617 - 0.1 * (42075.508) = -4265.1678$$

$$w[1] = 69.521 - 0.1 * (1725095.828) = -172440.0618$$

$$w[2] = 284.933 - 0.1 * (5806420.104) = -580357.0774$$

$$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -4265.1678 + -172440.0618 * 41 + -580357.0774 * 138$$

$$= -87163584.3828$$

$$\hat{y}^{(2)} = w[0] + w[1] * 42 + w[2] * 153 = -4265.1678 + -172440.0618 * 42 + -580357.0774 * 153$$

$$= -96041380.6056$$

$$\hat{y}^{(3)} = w[0] +$$

$$w[1] * 37 + w[2] * 151 = -4265.1678 + -172440.0618 * 37 + -580357.0774 * 151$$

$$= -94018466.1418 \quad (۷)$$

$$\hat{y}^{(4)} = w[0] + w[1] * 46 + w[2] * 133 = -4265.1678 + -172440.0618 * 46 + -580357.0774 * 133$$

$$= -85123999.30479999$$

$$MSE = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{8} \sum_{i=1}^{4} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 =$$

$$\frac{1}{8} \sum_{i=1}^{4} \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{8} \left[(-87163584.3828 - 37.99)^2 + (-96041380.6056 - 47.34)^2 \right.$$

$$\left. + (-94018466.1418 - 44.38)^2 + (-85123999.30479999 - 28.17)^2 \right]$$

$$= 4113379165155784.5$$

ج) به روزرسانی به شرح زیر است

$$J\left(w_0 \cdot w_1 \cdot w_2\right) = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$\frac{\partial J\left(w_0 \cdot w_1 \cdot w_2\right)}{\partial w_j} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)x_j^{(i)} \text{ □□□ } j = 1.2$$

$$\frac{\partial J\left(w_0 \cdot w_1 \cdot w_2\right)}{\partial w_0} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

$$\frac{\partial J\left(w_0 \cdot w_1 \cdot w_2\right)}{\partial w_0} = \frac{1}{4}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right) = \frac{1}{4}[(17.15 - 37.99) + (26.00 - 47.34) + (25.55 - 44.38)$$

$$+ (13.40 - 28.17)] = -18.945$$

$$\frac{\partial J\left(w_0 \cdot w_1 \cdot w_2\right)}{\partial w_1} = \frac{1}{4}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)x_1^{(i)} = \frac{1}{4}[(17.15 - 37.99) * 41 + (26.00 - 47.34) * 42$$

$$+ (25.55 - 44.38) * 37 + (13.40 - 28.17) * 46] = -781.7125$$

$$\frac{\partial J\left(w_0 \cdot w_1 \cdot w_2\right)}{\partial w_2} = \frac{1}{4}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)x_2^{(i)} = \frac{1}{4}[(17.15 - 37.99) * 138 + (26.00 - 47.34) * 153$$

$$+ (25.55 - 44.38) * 151 + (13.40 - 28.17) * 133] = -2737.17 \qquad (۸)$$

$$w[0] = -59.50 - 0.1 * (-18.945) = -57.6055$$

$$w[1] = -0.15 - 0.1 * (-781.7125) = 78.02125$$

$$w[2] = 0.60 - 0.1 * (-2737.17) = 274.317$$

$$\hat{y}^{(1)} = w[0] + w[1] * 41 + w[2] * 138 = -57.6055 + 78.02125 * 41 + 274.317 * 138 = 40997.01175$$

$$\hat{y}^{(2)} = w[0] + w[1] * 42 + w[2] * 153 = -57.6055 + 78.02125 * 42 + 274.317 * 153 = 45189.788$$

$$\hat{y}^{(3)} = w[0] + w[1] * 37 + w[2] * 151 = -57.6055 + 78.02125 * 37 + 274.317 * 151 = 44251.04775$$

$$\hat{y}^{(4)} = w[0] + w[1] * 46 + w[2] * 133 = -57.6055 + 78.02125 * 46 + 274.317 * 133 = 40015.5334$$

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 =$$

$$\frac{1}{8}\sum_{i=1}^{4}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{8}\left[(40997.01175 - 37.99)^2 + (45189.788 - 47.34)^2\right.$$

$$+ (44251.04775 - 44.38)^2 + (40015.533 - 28.17)^2 = 908587593.4252921$$

د) همانطور از مقادیر تابع هزینه معلوم است، در روش GD وزن‌ها بهتر به‌روزرسانی شده اند و مقدار تابع هزینه از روش SGD کمتر است. به این دلیل که روش SGD بیشتر تحت تاثیر نویز داده‌ها قرار گرفته و مسیر بهینه‌سازی پراعوجاج‌تری را طی می‌کند که منجر به واریانس زیاد در به‌روزرسانی می‌شود. درحالی که در روش GD اثر نویزی داده‌ها تا حدودی ازبین می‌رود. برای عملکرد بهتر SGD لازم است داده‌ها بسیار بیشتر باشند.

نکات تکمیلی

۱. لزومی به تایپ کردن سوالات تئوری نیست؛ ولی در صورتیکه پاسخ آنها به صورت تایپ‌شده تحویل داده شود، ۵ درصد نمره اضافه به شما تعلق میگیرد. در صورتیکه پاسخ‌های شما تایپ‌شده نیست، باید پاسخ‌ها خوانا و باکیفیت در قالب فایل pdf ارسال شوند.

۲. فرمت نامگذاری تکلیف ارسالی باید به صورت زیر باشد: HWX_Theory_LastName_StudentID که X شماره تکلیف LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.

۳. انجام این تکلیف به صورت تک نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می‌شود.

۴. برای تکالیف تئوری امکان ارسال با تاخیر وجود ندارد.

۵. در صورت وجود هر گونه ابهام و یا سوال می‌توانید سوالات خود را در گروه تلگرام بپرسید. هم چنین می‌توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.

آیدی‌ها:

@AlirezaT

@Yasinhmv