



برآورد تابع رگرسیون نیمه پارامتری با خطاهای همبسته بر اساس موجک انقباضی برای کاربردهای یادگیری نیمه نظارت شده

حیدر مختاری فریور^{۱*}، حبیب ایزدخواه^۲، بهزاد قلی پور^۳ و عبدالشکور تمندانی^۴

^۱ پژوهشگر، دانشگاه تبریز، تبریز، mokhtari@pgu.ac.ir

^۲ دانشیار گروه علوم کامپیوتر، دانشگاه تبریز، تبریز

^۳ پژوهشگر، دانشگاه خلیج فارس، بوشهر

^۴ مربی، دانشگاه ولایت، ایرانشهر

چکیده. آنالیز موجکی یکی از ابزارهای کارآمد در بسیاری از مباحث مهندسی است و علیرغم عمر کوتاه خود در حوزه های مختلف رشد گسترش یافته است. در این مقاله، ضمن معرفی تبدیل موجک، تابع رگرسیون نیمه پارامتری زمانی که داده ها به عنوان اطلاعات بدست آمده از یک سیگنال و همراه با خطاهای گوسین همبسته می باشد به روش موجک انقباضی برآورد می شود. سپس نرخ همگرایی برآوردگر در قالب چند لم و قضیه تعیین می شود. توابع رگرسیونی نیمه پارامتری نقش مهمی در تشخیص های نیمه نظارت شده دارد. در بخش آخر مطالب تئوری بدست آمده با استفاده از شبیه سازی مورد بررسی قرار می گیرد و متوسط میانگین مربعات خطای برآوردگر پیشنهادی با برآوردگر هسته ای مقایسه می شود. در پایان عملکرد برآوردگر پیشنهادی را در قالب یک مثال کاربردی مرتبط با میزان رشد دندان با توجه به مقدار ویتامین C دریافتی مورد بررسی قرار می گیرد. موجک انقباضی، رگرسیون نیمه پارامتری، توزیع گوسین، نرخ همگرایی، یادگیری نیمه نظارت شده
 طبقه بندی موضوعی [۲۰۱۰]: 62M45, 68T45, 68T05.

۱. مقدمه

رگرسیون و تحلیل رگرسیونی حوزه ای است که به بررسی ارتباط بین متغیر مستقل و متغیرهای وابسته می پردازد. رگرسیون با توجه به نوع تابعی که ارتباط بین متغیر مستقل و پاسخ را بیان می کند به دو دسته رگرسیون پارامتری و ناپارامتری تقسیم می شود. در رگرسیون پارامتری ارتباط بین متغیر مستقل X و متغیر پاسخ Y از طریق تابعی مانند g به صورت زیر تعیین می شود:

$$y_i = g(x_i|\beta) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2),$$

که در آن β بردار پارامترهای مجهول تابع g است و در صورت برآورد آن با بردار $\hat{\beta}$ تابع پارامتری g و در نتیجه مدل رگرسیونی به صورت $Y = g(X|\hat{\beta})$ برآورد می شود. در رگرسیون ناپارامتری ارتباط بین متغیر مستقل و متغیر پاسخ از طریق تابعی مجهول مانند f بیان می شود. رگرسیون ناپارامتری به خاطر پیش شرطهای اولیه کم از اهمیت خاصی برخوردار است.

یکی از ابزارهای مهم در برآورد تابع رگرسیونی موجک است. موجک ابزاری آنالیزی در پردازش سیگنال ها و مدل بندی ریاضی محسوب می شود و در ریاضیات کاربردی توانایی حل مسائل مربوطه را دارا می باشد. در صورتی که سیگنال، نمایشی از تابع زمانی باشد، موجک موقعیت بندی مناسب و کارایی را در هر دو زمینه مقیاسی و مکانی ارائه می دهد، چیزی که پیش از این برای تبدیل فوری امکان پذیر نبود. بطور کلی هر تجزیه و تحلیل مبتنی بر تبدیل فوری را می توان با استفاده از موجک ها

*ارائه دهنده

نیز مدل بندی کرد و اطلاعات زمانی و مکانی بیشتری بدست آورد. در سال‌های اخیر، نظریه موجک‌ها پیشرفت فراوانی در بسیاری از زمینه‌های علوم مختلف از جمله آمار داشته است. اولین بار موجک را داناها و جانستون [۹] (۱۹۹۲) در تحقیقات آماری بکار بردند. با افزایش توجه به موجک، کاربردهای بیشتری از این ابزار در علم آمار ارائه شده است که به عنوان برخی از کارهای اخیر در این زمینه می‌توان به واسرمن [۱۵] و همکاران (۲۰۰۵)، [۵] انتونیودیس (۲۰۰۷)، [۱۴] ویداکویک (۲۰۰۹)، [۱۳] نیسون (۲۰۱۱) و [۴] افشاری (۲۰۱۳) اشاره کرد.

در این مقاله ابتدا در بخش دوم، به معرفی موجک انقباضی، مقدار آستانه و نحوه استفاده از این روش در برآوردیابی و اغتشاش‌زدایی خواهیم پرداخت. در بخش سوم، مدل رگرسیون نیمه‌پارامتری را با فرض اینکه خطاها همبسته و دارای توزیع گوسین باشند، با استفاده از موجک انقباضی برآورد و نرخ همگرایی برآوردگر مطلوب در قالب چند لم و قضیه بیان می‌شود. در بخش چهارم، با استفاده از شبیه‌سازی، دقت روش پیشنهادی در مقایسه با روش برآوردیابی هسته‌ای مورد ارزیابی قرار می‌گیرد. به عنوان مثال عملی، مدل رگرسیون موجکی بر روی داده‌های مربوط به میزان رشد دندان برازش داده شده است.

۲. موجک انقباضی و روشهای آستانه‌گذاری

موجک‌ها ابزاری قابل انعطاف با محتوای غنی ریاضی هستند که در تجزیه و تحلیل یک تابع نوسانی از زمان و مکان بکار می‌روند. در تجزیه موجکی موجک اصلی (تابع یا سیگنال) بر روی دسته‌ای از توابع پایه که موجک نامیده می‌شوند تصویر شده و با استفاده از حاصل ضرب موج اصلی و تابع موجک تجزیه موجکی انجام می‌پذیرد. موجک تابعی با میانگین صفر است که بسط بر اساس انتقال و اتساع انجام می‌شود. اساس موجک از موجک پدر $\phi(\cdot)$ و موجک مادر $\psi(\cdot)$ تشکیل شده که به ترتیب روند کلی داده‌ها و انحرافات را نشان می‌دهند و به صورت زیر تعریف می‌شوند:

$$\psi(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{4}, \\ -1, & \frac{1}{4} \leq x < \frac{1}{2}, \\ 0, & O.W \end{cases} \quad \phi(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & O.W \end{cases}$$

برای $0 \leq j < \infty$ و $k = 0, 1, \dots, 2^j - 1$ انتقال‌های مقیاسی ϕ و ψ به صورت زیر تعریف می‌شوند.

$$\phi_{j,k}(x) = 2^{\frac{j}{2}} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k).$$

توابع ϕ و ψ و $\phi_{j,k}(x)$ و $\psi_{j,k}(x)$ که قطعه وار ثابت هستند می‌توانند در برآورد توابع استفاده شوند و برای هر j ثابت $\psi_{j,k}$ را موجک‌های نسل j ام گوئیم. اگر برای هر $j \in Z$ ، $V_j = \text{Span}\{\phi_{j,k}, k \in Z\}$ ، در این صورت V_j ‌ها دنباله‌ای از فضاها ی تودرتو به صورت $V_j \subset V_{j+1}$ خواهند بود. همچنین تصویر f روی فضای V_j بصورت زیر می‌باشد:

$$P_j f = \sum_{k \in Z} \alpha_{j,k} \phi_{j,k}(x), \quad \alpha_{j,k} = \int \phi_{j,k}(x) f(x) dx.$$

اگر W_j را فضای تولید شده از متمم متعامد V_j در V_{j+1} در نظر بگیریم در این صورت $W_j = \text{Span}\{\psi_{j,k}, k \in Z\}$ با فرض $f \in L^2(R)$ تقریب تابع f بر روی فضای V_j بصورت زیر خواهد بود:

$$(1) \quad \tilde{f} = \sum_{k \in Z} \alpha_{j_0,k} \phi_{j_0,k}(x) + \sum_{j > j_0} \sum_{k \in Z} \beta_{j,k} \psi_{j,k} = P_{j_0} f + \sum_{j > j_0} D_j f,$$

که در آن $\beta_{j,k} = \int \psi_{j,k}(x) f(x) dx$ و فوق D_j تصویر f بر W_j است (برای اطلاعات بیشتر به دوبیشی (۱۹۹۲) مراجعه شود) [۸].

تبدیل موجک ابزاری برای تجزیه کردن توابع، عملگرها و داده‌ها به اجزای با فرکانسهای مختلف است که اجازه مطالعه هر کدام از اجزا را بطور جداگانه می‌دهد. در تبدیل موجک گسسته، سیگنال از یک سری فیلترهای بالاگذر برای فرکانس‌های بالا و یک سری فیلترهای پایین‌گذار برای آنالیز فرکانس‌های پایین، عبور داده شده و سیگنال به دو بخش تقسیم می‌شود: بخش حاصل از عبور سیگنال بالاگذر که شامل اطلاعات فرکانس بالا (از جمله اغتشاش) است و جزئیات نام دارد، بخش حاصل از عبور سیگنال از فیلتر پایین‌گذر که شامل اطلاعات فرکانس پایین و در برگیرنده مشخصات هویتی سیگنال بوده و کلیات نامیده می‌شود. در نهایت، گروهی از سیگنال‌ها را خواهیم داشت که همان سیگنال اولیه را نشان می‌دهند اما هر گروه سیگنال

به باند فرکانسی متفاوتی مربوط است.

اغتشاش به معنای آلودگی صوتی و یا سیگنالی ناخواسته است که شکل سیگنالها را تغییر می‌دهد و باعث بروز اختلال می‌شود. اغتشاش اغلب دارای اجزاء فرکانسی نیز بوده و تجزیه و تحلیل فرکانسی شبیه به سیگنال دارد. روش‌های مختلفی برای اغتشاش زدایی از سیگنال‌هایی که به همراه اغتشاش مشاهده می‌شوند وجود دارد. داناها و جانستون (۱۹۹۴) روش آستانه‌ای را برای حذف اغتشاشات ارائه دادند که به روش موجک انقباضی مشهور شد. [۱۰] دلایل زیادی برای استفاده از این روش وجود دارد. از جمله می‌توان به سادگی استفاده از آن، کاربردی و سریع بودن، قابلیت استفاده در مسائل مختلف اشاره کرد. از آنجا که برآوردگر موجکی انقباضی دارای ریسکی نزدیک به ریسک برآوردگر کم بیشینه است، دارای معیار میانگین توان دوم خطای کمینه می‌باشد و همچنین در تجزیه سیگنال‌های ناهمگن به خوبی عمل می‌کند. روش موجک انقباضی مبتنی بر دو پارامتر اساسی مقدار آستانه و تابع حذف است. با مشخص شدن مقدار آستانه و تابع حذف، مراحل الگوریتم موجک انقباضی بصورت زیر خواهد بود:

(۱) ابتدا از مشاهدات اغتشاشی تبدیل موجک گسسته گرفته می‌شود. به عبارت دیگر، با فرض اینکه ماتریس تبدیل موجکی گسسته W ، و مدل مشاهدات به صورت $y = f + \epsilon$ باشد ابتدا ضرایب موجکی را بدست می‌آوریم.

$$Wy = Wf + W\epsilon \Rightarrow d^* = d + \epsilon,$$

(۲) با استفاده از مقدار آستانه، ضرایب موجکی به دو دسته ضرایب پر اهمیت و ضرایب کم اهمیت دسته‌بندی می‌شوند به اینصورت که اگر ضریب موجکی بیشتر از مقدار آستانه باشد، متعلق به دسته ضرایب پر اهمیت است و در غیر اینصورت جزو دسته ضرایب کم اهمیت خواهد بود. سپس ضرایب کم اهمیت حذف شده و ضرایب پر اهمیت با توجه به تابع حذف، اصلاح می‌شوند.

(۳) سیگنال حاصل را با استفاده از تبدیل موجک معکوس بازسازی می‌کنیم. در مرحله دوم، می‌توان از تابع حذف سخت و یا نرم استفاده کرد. در تابع حذف سخت، ضرایب کمتر از مقدار آستانه برابر صفر و باقی ضرایب بدون تغییر باقی می‌مانند. به عبارت دیگر آستانه سخت برابر است با:

$$\eta_{hard}(x) = xI_{\{|x| < \lambda\}}$$

در تابع حذف نرم، ضرایب کمتر از مقدار آستانه برابر صفر و باقی ضرایب به اندازه مقدار آستانه به صفر نزدیک می‌شوند. به عبارت دیگر آستانه نرم برابر است با:

$$\eta_{soft}(x) = (x + \lambda)I_{\{x < \lambda\}} + (x - \lambda)I_{\{x > \lambda\}}$$

همچنین روش‌های مختلفی برای تعیین آستانه وجود دارد که از آن جمله می‌توان به آستانه جهانی، اعتبار متقابل و قطعی اشاره نمود. در این مقاله برای برآورد تابع از مقدار آستانه جهانی $\lambda = \sigma \sqrt{2 \log(n)}$ استفاده شد که در آن σ نشان دهنده انحراف معیار اغتشاشات و n تعداد مشاهدات است.

۳. برآورد موجک انقباضی مدل رگرسیون نیمه پارامتری

در زمینه برآورد تابع رگرسیون بر اساس موجکها تحقیقات فراوانی انجام شده است که می‌توان به هال و پاتیل [۱۲] (۱۹۹۶)، [۶] بیگوت و ساپاستیناس (۲۰۰۱)، واسرمن و همکاران [۱۵] (۲۰۰۵) [۱۱] و دوستی و همکاران (۲۰۰۸) اشاره کرد که روش برآورد موجک انقباضی در رگرسیون ناپارامتری را تحت شرایط مختلف مورد بررسی قرار دادند. در اکثر روش‌های رگرسیون ناپارامتری تابع مورد نظر پیوسته و همواره مشتق پذیر در نظر گرفته می‌شود در حالی که در برآورد موجکی، نیازی به مشتق پذیری تابع رگرسیون نیست. در این بخش ابتدا رگرسیون نیمه پارامتری را معرفی نموده و برآورد موجکی انقباضی این مدل را محاسبه و در انتها نرخ همگرایی برآوردگر بررسی می‌شود.

تعریف ۱.۳. اگر $f^{(m)}$ نشان دهنده m امین مشتق تابع f و $[\alpha]$ نیز بزرگترین عدد صحیح کوچکتر مساوی α باشد. آنگاه برای هر مقدار حقیقی α ، تابع f متعلق به کلاس هولدر $\Lambda^\alpha(M)$ روی $[0, 1]$ است هرگاه:

۱. اگر $0 \leq \alpha \leq 1$ ، $|f(x) - f(y)| \leq M|x - y|^\alpha$
۲. اگر $\alpha > 1$ ، $|f^{([\alpha])}(x) - f^{([\alpha])}(y)| \leq M|x - y|^{\alpha - [\alpha]}$ و $|f^{(1)}(x)| \leq M$

تعریف ۲.۳. فرض کنید $((t_1, y(t_1)), \dots, (t_n, y(t_n)))$ یک نمونه تصادفی از مدل زیر باشد:

$$(2) \quad y_i = f(t_i) + e_i, \quad t_i \stackrel{iid}{\sim} U[0, 1], \quad e_i \stackrel{iid}{\sim} N(0, \sigma^2),$$

که $0 \leq t_{(1)} \leq \dots \leq t_{(n)} \leq 1$ نشان‌دهنده آماره‌های ترتیبی t_1, \dots, t_n باشند بطوریکه در شرایط زیر صدق کنند.

$$\text{var}(t_{(i)}) \leq \frac{1}{n}, \quad |E(t_{(i)} - \frac{i}{n})| \leq \frac{1}{\sqrt{n}}, \quad i = 1, \dots, n,$$

$$\text{Cov}(e(t_{(i)}), e(t_{(j)})) = \gamma(|i - j|), \quad \lim_{n \rightarrow \infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| < \infty,$$

در این صورت مدل (۲) را مدل رگرسیون نیمه پارامتری می‌نامند.

با استفاده از رابطه (۱)، برآوردگر موجکی تابع f بصورت زیر خواهد بود:

$$(5) \quad \hat{f}(t) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0, k} \phi_{j_0, k}(t) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j, k} \psi_{j, k}(t).$$

که در آن $\hat{d}_{j, k}$ ها ضرایب موجکی و $\hat{c}_{j_0, k}$ ها ضرایب مقیاسی بدست آمده از تبدیل موجک گسسته هستند و J' بزرگترین عدد صحیحی است که برای مقدار ثابت $K > 0$ در نامساوی $2^{J'} \leq K \sqrt{\frac{n}{\log(n)}}$ صدق کند. در عمل معمولاً ثابت k را طوری انتخاب می‌شود که $J' \geq J$ باشد و ضرایب آستانه روی همه سطوح ابتدایی در سطح j_0 انجام می‌شود.

قضیه ۳.۳. اگر شرایط (۳) و (۴) برای مدل (۲) برقرار باشد و $|f(x) - f(y)| \leq M|x - y|^{s(\alpha)}$ بطوری که $s(\alpha) = \min(\alpha, 1)$ ، آنگاه برای هر $f \in \Lambda^\alpha(M)$ ، خطای برآورد با استفاده از رابطه (۵) از مرتبه $O(n^{-s(\alpha)})$ است. اثبات. با استفاده از رابطه جنسن داریم:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n E \left((f(t_{(i)}) - f(t)) \right)^2 &\leq \frac{M}{n} \sum_{i=1}^n E \left((t_{(i)} - t)^{2s(\alpha)} \right) \\ &\leq \frac{M}{n} \sum_{i=1}^n \left[E \left((t_{(i)} - t)^2 \right) \right]^{s(\alpha)} \\ &\leq \frac{M}{n} \sum_{i=1}^n \left[\frac{1}{n} + \left(\frac{1}{\sqrt{n}} \right)^2 \right]^{s(\alpha)} = O(n^{-s(\alpha)}). \end{aligned}$$

□

در ادامه چهار لم ارائه می‌شود که در اثبات قضیه اصلی این بخش مورد استفاده قرار می‌گیرند. برای سادگی در نوشتن، در اثبات لم‌ها و قضایای مربوطه، از ضرایب C_i به عنوان ضرایب ثابت استفاده شده است.

لم ۴.۳. فرض کنید $f \in \Lambda^\alpha(M)$ و موجک مادر ψ دارای r گشتاور صفر باشد بطوری که $r \geq \alpha$. آنگاه:

$$\sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} \hat{d}_{j, k}^2 \leq C_r \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+\alpha}}.$$

لم ۵.۳. فرض کنید $f \in \Lambda^\alpha(M)$ و موجک مادر ψ دارای r گشتاور صفر باشد بطوری که $r \geq \alpha$. همچنین فرض کنید برای هر متغیر تصادفی Y ، $E_1(Y) = E(Y|t_1, t_2, \dots, t_n)$. آنگاه:

$$\sum_{k=0}^{2^{j_0}-1} E \left((\hat{c}_{j_0, k} - c_{j_0, k}) \right)^2 \leq C_r 2^{2j_0} n^{-1} + 2 \sum_{k=0}^{2^{j_0}-1} \left(\hat{a}_{j_0, k}^2 + E(\hat{b}_{j_0, k}^2) \right).$$

لم ۶.۳. : تحت شرایط لم (۱) و لم (۲) رابطه زیر برقرار است:

$$E \left(\left(\hat{d}_{j,k} - d_{j,k} \right)^2 \right) \leq \wedge d_{j,k}^2 \wedge \epsilon \frac{2^j C_f}{n} \log n$$

$$+ 2^0 a_{j,k}^2 + 2^0 E(b_{j,k}^2) + \frac{2^{j+1} C_f}{n^2}.$$

بطوریکه $x \wedge y = \min(x, y)$.

لم ۷.۳. : تحت شرایط لم ۳ رابطه زیر برقرار است:

$$\sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} E \left(\left(\hat{d}_{j,k} - d_{j,k} \right)^2 \right) \leq C_5 \left(\frac{\log n}{n} \right)^{\frac{2\alpha}{2+\alpha}}$$

$$+ 2^0 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} a_{j,k}^2 + E(b_{j,k}^2)$$

قضیه ۸.۳. فرض کنید شرایط (۳) و (۴) در مدل (۲) برقرار باشد. همچنین فرض کنید موجک مادر ψ ، دارای r گشتاور متمایل به صفر و محمل فشرده است. آنگاه برآوردگر \hat{f} تعریف شده در رابطه (۵) تحت یک عامل لگاریتمی دارای نرخ همگرایی بهینه در فضای هولدر $\Lambda^\alpha(M)$ است. به عبارت دیگر برای هر $M \in (0, \infty)$ ، خواهیم داشت:

$$\sup_{f \in \Lambda^\alpha(M)} E \|\hat{f} - f\|_2^2 \leq C \left(\frac{\log n}{n} \right)^{\alpha/(1+\alpha)}$$

و

$$\sup_{f \in \Lambda^\alpha(M)} \frac{1}{n} \sum E \|\hat{f} - f\|_2^2 \leq C \left(\frac{\log n}{n} \right)^{\alpha/(1+\alpha)}$$

اثبات.

$$E \left(\|\hat{f} - f\|_2^2 \right) = E \left(\int_0^1 \left[\hat{f}(x) - f(x) \right]^2 dx \right)$$

$$= E \left(\int_0^1 \left[\sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k} \psi_{j,k}(x) \right. \right.$$

$$\left. - \sum_{k=0}^{2^{j_0}-1} c_{j_0,k} \phi_{j_0,k}(x) - \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right. \left. - \sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right]^2 dx \right)$$

$$= E \left(\int_0^1 \left[\sum_{k=0}^{2^{j_0}-1} (\hat{c}_{j_0,k} - c_{j_0,k}) \phi_{j_0,k}(x) \right. \right.$$

$$\left. + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} (\hat{d}_{j,k} - d_{j,k}) \psi_{j,k}(x) - \sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right]^2 dx \right)$$

جدول ۱. مقدار متوسط میانگین مربع خطا و انحراف استاندارد میانگین مربع خطاها برای توابع هدف

$$AMSE = \frac{1}{100} \sum_{j=1}^{100} \sum_{i=1}^n \frac{(f(x_i) - \hat{f}(x_i^j))^2}{n}$$

بر اساس ۱۰۰ بار تکرار

۱۰۲۴		۵۱۲		۲۵۶		n	تابع هدف
SD	AMSE	SD	AMSE	Sd	AMSE	برآوردگر	
۰۴۱/۰	۰۷۴/۰	۰۸۵/۰	۰۹۷۶/۰	۱۲۶/۰	۰۳/۱	موجکی	تابع سینوس
۰۵۸/۰	۰۹۷۱/۰	۰۹۳/۰	۰۹۸۵/۰	۱۲۸/۰	۰۹۷۴/۰	هستهای	
۰۱۸/۰	۰۱۷۶/۰	۰۲۶/۰	۰۱۸۲/۰	۰۴۳/۰	۰۲۱۷/۰	موجکی	تابع بلوکی
۰۲۱/۰	۰۱۹۰/۰	۰۲۹/۰	۰۱۹۱/۰	۰۴۶/۰	۰۱۸۸/۰	هستهای	

به دلیل متعامد یکدیگر بودن پایه موجک، عبارت فوق را می‌توان به صورت زیر نوشت:

$$E(\|\hat{f} - f\|_2^2) = \sum_{k=0}^{2^j-1} E\left((\hat{c}_{j,k} - c_{j,k})^2\right) + \sum_{j=J'}^{J'-1} \sum_{k=0}^{2^j-1} E(\hat{d}_{j,k} - d_{j,k})^2 + \sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}^2$$

با استفاده از لم های ۱ و ۲ و ۴، داریم:

$$E(\|\hat{f} - f\|_2^2) \leq \frac{C_2 2^{2j}}{n} + \frac{C_3}{n^{2(1/2+s(\alpha))}} + \frac{C_4}{n^{s(\alpha)}} + C_5 \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)} + C_6 \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)}$$

همچنین برای هر $z \geq 0$ که در رابطه $C_2 2^{-j(2+2\alpha)} \leq \frac{\epsilon C_4}{\lambda n} \log n$ صدق می‌کند، J_1 ای وجود دارد بطوریکه، $j \geq J_1$ و $\frac{2C_4 2^{2J'}}{n^2} \leq 2C_4 K \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}}$ و $\frac{\epsilon C_4 \log n}{n} 2^{2J_1} \leq C_9 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}}$ ، $\frac{d_{j,k}^2}{2^j} \leq C_7 2^{-j(2+2\alpha)} \leq \frac{\epsilon C_4}{\lambda n} \log n$ از طرفی، $\frac{C_2 2^{2j}}{n} \leq C_{12} \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)}$ ، $\frac{C_3}{n^{2(1/2+s(\alpha))}} \leq C_{10} \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)}$ ، برای هر $\alpha \in \{0, 1\}$ می‌توان نوشت:

$$C_{11}/n^{s(\alpha)} \leq C_{11} \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)}$$

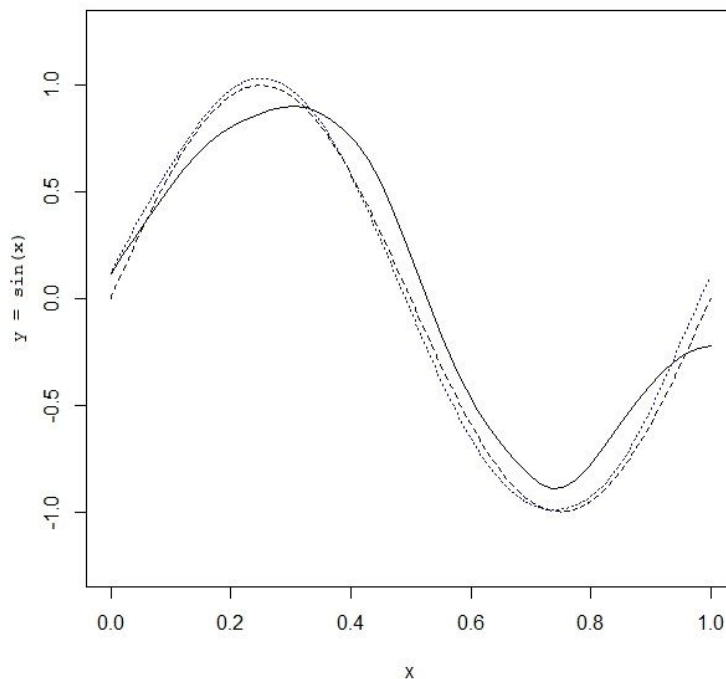
بنابراین اثبات به صورت زیر کامل می‌شود.

$$E(\|\hat{f} - f\|_2^2) \leq C \left(\frac{\log n}{n}\right)^{\alpha/(1+\alpha)}$$

□

۴. مطالعه شبیه‌سازی

در این بخش، نتایج بدست آمده از بخش تئوری را با استفاده از شبیه‌سازی مورد بررسی قرار می‌گیرد و نتایج بدست آمده را با نتایج حاصل از روش برآورد رگرسیون ناپارامتری هسته‌ای با استفاده از معیار متوسط میانگین مربعات خطاهای آزمایشات (AMSE) مقایسه می‌شود. همچنین، از تابع سینوس و تابع بلوکی در بازه واحد به عنوان توابع هدف استفاده شده است.



شکل ۱. شکل ۱- برآورد تابع سینوس با استفاده از دو روش موجکی و هسته‌ای. نمودار نقطه چین و خط پیوسته به ترتیب نشان دهنده برآوردگر موجکی و هسته‌ای هستند. تابع هدف با خط بریده نشان داده شده است.

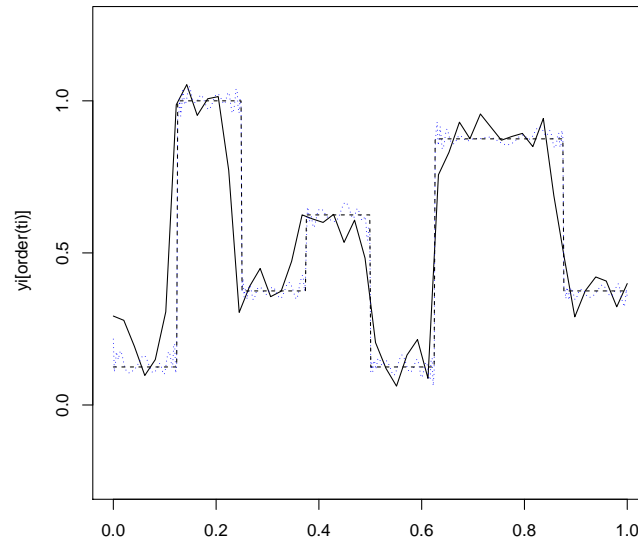
هر آزمایش شامل ۲۵۶، ۵۱۲ و ۱۰۲۴ مشاهده از تابع هدف با شرایط مدل رگرسیون نیمه پارامتری است. در هر آزمایش، علاوه بر معیار AMSE، مقدار انحراف معیار برآوردگرها نیز محاسبه شده است. شبیه‌سازی با استفاده از نرم‌افزار R و مقدار آستانه جهانی و تابع حذف سخت انجام گرفته است.

همانطور که در جدول ۱ مشاهده می‌شود، همواره برآوردگر موجکی مقدار AMSE کمتری نسبت به برآوردگر هسته‌ای دارد. همچنین، با افزایش تعداد مشاهدات مدل نمونه‌گیری، دقت برآوردگر موجکی نسبت به برآوردگر هسته‌ای بیشتر می‌شود و میزان خطای برآوردگر موجکی با سرعت بیشتری به سمت صفر میل می‌کند. علاوه بر این، برآوردگر موجکی همواره خطای استاندارد کمتری نسبت به برآوردگر هسته‌ای دارد. برای رسم شکل ۱ و ۲، از یک نمونه به اندازه ۵۱۲ استفاده شده است. همانطور که در شکل ۲ دیده می‌شود، برآوردگر موجکی، نقاط نک‌تیز با پرش را نسبت به برآوردگر هسته‌ای بهتر برآورد می‌کند.

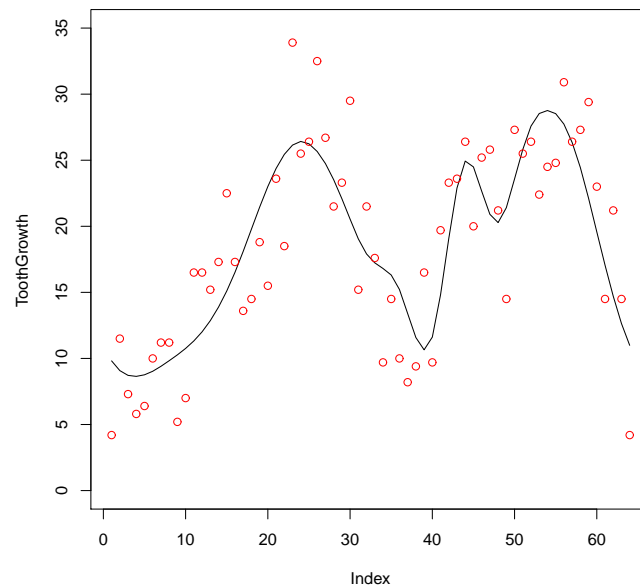
مثال ۱۰۴: به عنوان مثال تجربی، از داده‌های ToothGrowth که بر روی بسته datasets نرم‌افزار R قرار دارد مورد بررسی قرار می‌گیرد. این چارچوب اطلاعاتی اثر ویتامین C را بر روی میزان رشد طول دندان‌ها نشان می‌دهد و شامل ۶۰ داده از این دسته است. در روش موجکی تعداد داده‌ها باید ضریبی از ۲ باشد. برای برطرف کردن این مشکل، یک نمونه به حجم چهار بدون جایگذاری به تصادف از بین داده‌های موجود انتخاب و به داده‌های اولیه اضافه می‌شود. نمودار زیر مدل حاصل از روش رگرسیون ناپارامتری موجکی را نشان می‌دهد.

۵. بحث و نتیجه‌گیری

در این مقاله برآوردگر موجکی انقباضی رگرسیون نیمه‌پارامتری مورد بررسی قرار گرفت و نرخ همگرایی آن محاسبه گردید. به منظور ارزیابی توانایی روش پیشنهادی، در بخش شبیه‌سازی از دو تابع سینوسی و تابع بلوکی به عنوان توابع هدف استفاده شد و متوسط میانگین مربعات خطاهای بدست آمده از این روش با متوسط میانگین مربعات خطاهای روش هسته‌ای مقایسه گردید. نتایج بدست آمده نشان می‌دهد که روش موجکی انقباضی همواره مقدار AMSE کمتری نسبت به روش هسته‌ای



شکل ۲. شکل ۲- برآورد تابع بلوکی با استفاده از دگر روش موجکی و هسته‌ای. نمودار نقطه چین و خط پیوسته به ترتیب نشان دهنده برآوردگر موجکی هستند. تابع هدف با خط بریده نشان داده شده است.



شکل ۳. شکل ۴- استفاده از آستانه سخت مدل رگرسیون موجکی انقباضی، بر داده‌های مربوط به میزان رشد دندان برازش داده شد

دارد. همچنین، روش موجکی انقباضی، بنا به خاصیت پایه‌های موجکی و مقیاسی، نقاط تیز و جهشی را بهتر از سایر روش‌ها برآورد می‌کند و دارای نرخ همگرایی قابل قبول می‌باشد. به عنوان مثال کاربردی با استفاده از آستانه سخت مدل رگرسیون موجکی انقباضی، بر داده‌های مربوط به میزان رشد دندان برازش داده شد.



شکل ۴. برآورد موجکی انقباضی رگرسیون نیمه پارامتری

مراجع

- [1] حیدر مختاری فریور، محمود افشاری، (۱۳۹۴). برآورد تابع رگرسیونی با نقطه مشخص تغییر کننده با استفاده از موجک های کرانه ای. (هشتمین همایش ملی تخصصی آمار-۶-۷ اسفند ۱۳۹۴ دانشگاه پیام نور تهران- مجموعه مقالات ص ۸۴۷-۸۵۸)
- [2] حیدر مختاری فریور، محمود افشاری، (۱۳۹۴). کاربرد داده کاوی در برآورد رگرسیونی چند متغیره توزیع شده و یادگیری ماشین بر اساس موجک ها. (هشتمین همایش ملی تخصصی آمار-۶-۷ اسفند ۱۳۹۴ دانشگاه پیام نور تهران- مجموعه مقالات ص ۸۵۹-۸۷۵)
- [3] حیدر مختاری فریور، عبدالمجید مصلح، محمود افشاری، (۱۳۹۵). بررسی الگوی های تحلیل اطلاعات بر روی داده های توزیع شده با مشخصه بلادرنگ. همایش ملی روش های تحلیل اطلاعات- ۱۶ خرداد ۱۳۹۵، دانشگاه اطلاعات و امنیت ملی (وزارت اطلاعات). مجموعه مقالات
- [4] Afshari M, (2013). A fast wavelet algorithm for analyzing of signal processing and empirical distribution of Wavelet coefficients with numerical example and simulation. *Communication of statistics-Theory and methods*. 42: 4156-4169.
- [5] Antoniadis, A., 2007. Wavwelet methods in satatistics, some recent developments and their applications. *Statistical Surveys*. 1: 16-55
- [6] Bigot, A. J., and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6, pp-1.
- [7] Chui, C.K. (1999). *An introduction to wavelets*. Academic Press, Boston.
- [8] Daubechies, I. (1992). *Ten lectures in Wevelets*. CBMS-NSF regional conferences series in applied mathematics, SFAM, philadelphia.
- [9] Donoha, D.L., Johnstone, I.M. (1992). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 81:425-455.
- [10] Donoha, D.L., Johnstone, I.M. (1992). Minimax estimation by wavelet shrinkage. *Technical Report, Department of Statistics, Stanford University*.
- [11] Doosti, H., Afshari, M., Niroomand, H.A. (2008). Wavelets for nonparametric stochastic regression with mixing stochastic process. *Communication of statistics-Theory and methods*. Canada, Volume 37:373-385.
- [12] Hall, P., & Patil, P. (1996). On the choice of smoothing parameter, threshold and truncation in non-parametric regression by non-linear wavelet methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 361-377.
- [13] Nason G, (2011). *Wavelet Methods in Statistic with R*. Springer.
- [14] Vidakovic, B. (2009). *Statistical Modeling By Wavelets*. Wiley, NewYork.
- [15] Wasserman L, Genovese R, (2005) Confidence Sets for Nonparametric Wavelet Regression. *Institute of Mathematical Statistics*. 3: 698-729.