روش دینامیکی در حل معادلات رگرسیون خطی و غیر خطی

۱- مقدمه

هوش مصنوعی از تکنیک‌های متعددی برای بهبود عملکرد مدل‌های ماشینی و آموزش آنها استفاده می‌کند. الگوریتم‌های بهینه‌سازی مانند گرادیان کاهشی، نقطه عطفی در این فرآیند به‌شمار می‌آیند. با این حال، علمای دانشگاه‌ها و محققان صنعتی همچنان در جستجوی روش‌های بهتر و کارآمدتر برای بهبود گرادیان کاهشی و افزایش سرعت آموزش مدل‌های ماشینی هستند.

در این مقاله، یک روش جدید و نوآورانه مبتنی بر معادلات دینامیکی برای بهبود فرایند حل مسائله معرفی می‌شود. ما به بررسی ریاضیاتی و فیزیکی این روش می‌پردازیم و نشان می‌دهیم که چگونه این معادلات دینامیکی می‌توانند بهبود قابل ملاحظه‌ای در عملکرد گرادیان کاهشی داشته باشند. همچنین، مزایا و معایب این روش نسبت به روش‌های سنتی گرادیان کاهشی مورد بررسی قرار می‌گیرد.

در پایان، ما از انجام آزمایش‌های عملی و مقایسه نتایج با روش‌های متداول گرادیان کاهشی به منظور تایید اثربخشی این روش جدید گزارش می‌کنیم و پیشنهاداتی برای کاربردهای آینده و تحقیقات جدید ارائه می‌دهیم.

۲- رابطه سازی روش دینامیکی

فرض کنید مدلی به صورت رابطه (۱) داریم، می‌خواهیم ماتریس وزن های تابع را که به صورت رابطه (۲) است به ازای مقادیر مختلف ویژگی‌ها (۳) که تشکیل ماتریسی مشابه ماتریس (۵) را می‌دهند بیابیم.

$$P = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 \quad (۱)$$

$$W = \{\alpha_0 \quad \alpha_1 \quad \alpha_2\} \quad (۲)$$

$$X = \{1 \quad x_1 \quad x_2\} \quad (۳)$$

$$P = \{X\}\{W\}^T \quad (۴)$$

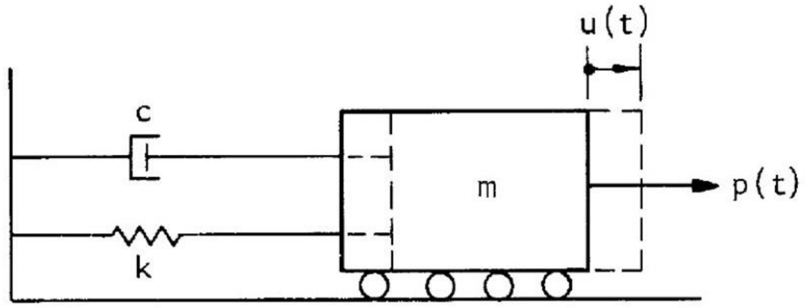
مقادیر مختلف ویژگی‌ها را در بردار (۳) جایگزین می‌کنیم و ماتریس (۵) تشکیل می‌شود.

$$S = \begin{bmatrix} 1 & 1 & & 1 & 1 \\ a_1 & a_3 & \dots & a_5 & a_7 \\ a_2 & a_4 & & a_6 & a_8 \end{bmatrix} \quad (۵)$$

لذا رابطه (۱) به صورت رابطه (۶) در خواهد آمد.

$$P = S^T W \quad (۶)$$

فرض کنید جسمی مشابه جسم شکل (۱) داریم، در معادله دینامیکی این جسم در حال نوسان به صورت رابطه (۷) است.



شکل (۱) – نمایش جسمی در حال نوسان

$$P = M\ddot{X} + C\dot{X} + S^T X \qquad (۷)$$

در رابطه بالا M مقدار جرم جسم مورد نظر است و C میرایی سیستم و S مقدار سختی فنر است و X مقادیر جابه جایی جسم را در لحظه t نشان می‌دهد. با در نظر گرفتن جرم ساختگی و میرایی ساختگی برای سیستم معادلات رابطه (۶) رابطه (۷) معادل حالت دینامیکی سیستم مورد نظر خواهد بود. لذا با حل مقدار X مقادیر وزن ما در معادله (۶) بدست خواهند آمد.

حالا اگر رابطه را به صورت زیر مرتب کنیم می‌توانیم مقدار خطا را برای یک روش نموی در بدست آوردن وزن‌ها ارزیابی کنیم.

$$R = P - S^T X = M\ddot{X} + C\dot{X} \qquad (۸)$$

R در رابطه بالا میزان خطا را نمایش می‌دهد.

تغییرات خطا را در دو مرحله متوالی نسبت به تغییر X به صورت زیر در نظر می‌گیریم.

$$\frac{R_{n+1}\text{-}R_n}{\delta = \Delta \mathrm{x}} = \lambda \qquad (۹)$$

و در مرحله n+1 داریم.

$$R_{n+1} = M\ddot{X}_{n+1} + C\dot{X}_{n+1} \qquad (۱۰)$$

حال تابع X را به صورت زیر بیان می‌کنیم.

$$X(t) = A * \cos{(\omega t)} \qquad (۱۱)$$

لذا معادله (۱۰) به صورت زیر در می‌آید.

$$R_{n+1} = -MA\,\omega^2\cos(\omega t) - cMA\omega\sin(\omega t) \qquad (۱۲)$$

در معادله بالا میرایی به صورت ضریبی از جرم در نظر گرفته شده است. با جایگذاری رابطه اخیر در رابطه (۹) رابطه (۱۳) حاصل می‌شود.

$$\frac{-MA\,\omega^2\cos(\omega t) - cMA\omega\sin(\omega t) - R}{\delta} = \lambda \tag{۱۳}$$

مقدار δ مطابق رابطه (۱۴) است.

$$\delta = -\tau A\omega\sin(\omega t) \tag{۱۴}$$

با جایگذاری رابطه (۱۴) در (۱۳) و مرتب کردن آن برای R رابطه (۱۵) حاصل می‌شود.

$$R = -MA\,\omega^2\cos(\omega t) - cMA\omega\sin(\omega t) + \lambda\tau A\omega\sin(\omega t) \tag{۱۵}$$

حال از طرفین نسبت به زمان مشتق می‌گیریم تا مقدار λ را برای بیشترین مقدار خطا بدست آوریم. پس از انجام عملیات مشتق مقدار λ برابر رابطه (۱۶) خواهد شد.

$$\lambda = \frac{M(\cos(\omega t)\,c - \omega\sin(\omega t))}{\cos(\omega t)\,\tau} \tag{۱۶}$$

با جایگذاری رابطه اخیر در رابطه (۱۵) و پس از ساده سازی آن رابطه (۱۷) به دست می‌آید.

$$A = -\frac{R}{\mathrm{M}\omega^2}\cos(\omega t) = -\frac{R}{S^T}\cos(\omega t) \tag{۱۷}$$

با در نظر گرفتن رابطه (۱۷) که مقدار اختلاف دامنه نوسان فنر در دو مرحله متوالی است و جایگذاری آن در رابطه (۱۸) مقدار جابه‌جایی برای مرحله (n+1) بدست می‌آید.

$$x_{n+1} = x_n + \eta A\cos(\omega t) \tag{۱۸}$$

η در این رابطه نرخ یادگیری است. جرم ساختگی سیستم برابر نصف مقدار سختی فنر در نظر گرفته می‌شود لذا به صورت رابطه (۱۹) بدست می‌آید.

$$\mathrm{M} = \frac{\mathrm{S}}{2} \tag{۱۹}$$

مقدار میرایی (ω) برای ماتریس مربعی S به صورت رابطه (۲۰) است.

$$\omega = \sqrt{\frac{X^T S X}{X^T M X}} \tag{۲۰}$$

اگر ماتریس‌ها غیر مربعی باشند مقدار میرایی برابر رابطه (۲۱) است.

$$\omega = \sqrt{\frac{S}{M}} = \sqrt{2} \tag{۲۱}$$

در نهایت تابع هزینه برابر میانگین مربع خطاها در هر مرحله خواهد بود.

$$cost = \frac{R^T.R}{n} \tag{۲۲}$$

در این رابطه n تعداد ورودی‌ها یا تعداد اعضای بردار(یا ماتریس) R است

در نظر داشته باشید چنانچه تعداد وزن‌ها برای چند معادله را بخواهیم با یک مقدار ویژگی بیابیم ماتریس وزن‌ها از حالت برداری به ماتریسی در خواهد آمد، لذا n برابر تعداد ردیف‌های ماتریس R خواهد بود. مقدار X اولیه با توجه به اینکه شتاب و سرعت در لحظه اولیه برابر صفر است، مطابق رابطه زیر است.

$$x_0 = (S^T)^{-1}\, P \tag{۲۳}$$

۳- نمونه عددی

به منظور نشان دادن عملکرد روش، چند نمونه عددی در مقایسه با چند روش معمول مثل گرادیان کاهشی و روش Adam مورد بررسی قرار می‌گیرد و نتایج آن ارائه می‌گردد. کد برنامه با زبان پایتون نوشته شده است.

مجموعه داده‌های ورودی و هدف:

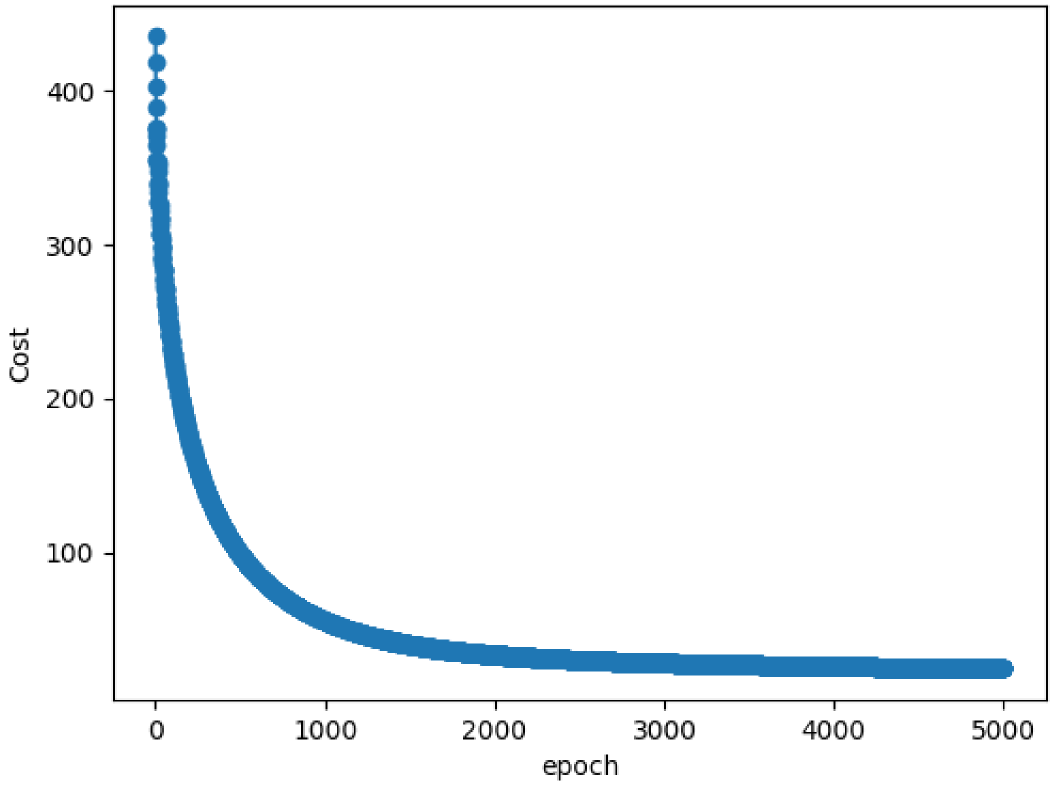
Data point = [(1, 2), (2, 9), (3, 16), (4, 25), (5, 36)]

۴- مدل و نتایج

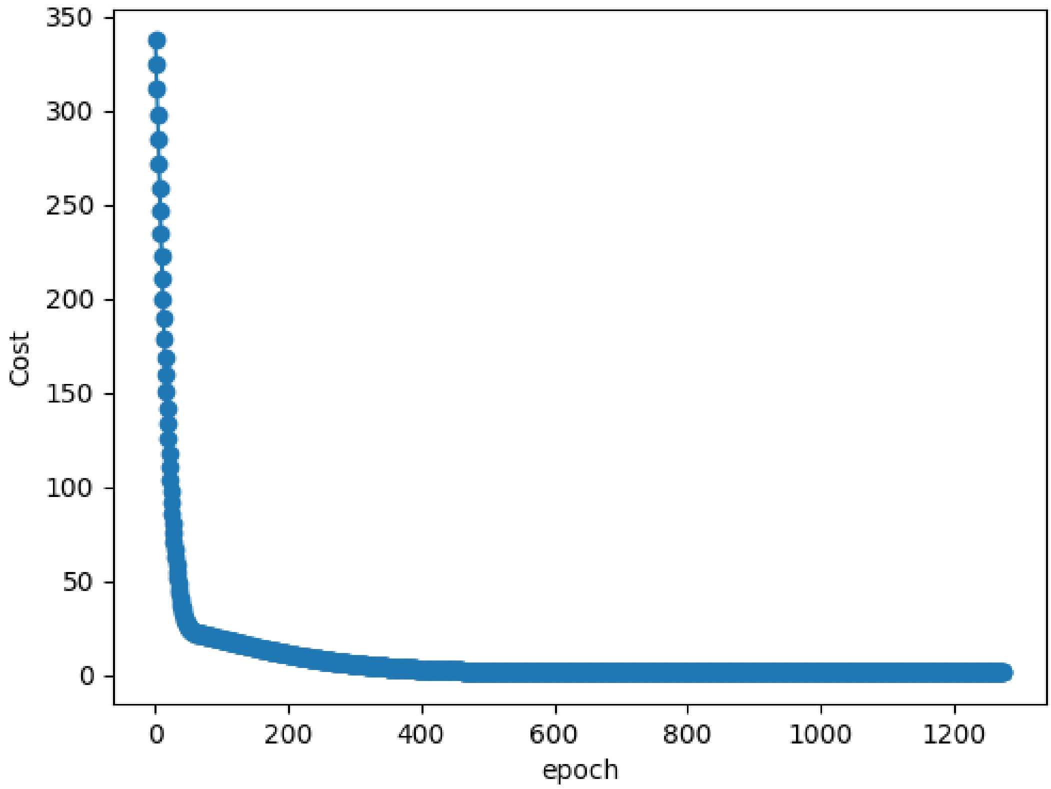
y = a*x+b

نرخ آموزش در روش‌های مبتنی بر گرادیان ۰.۱ در نظر گرفته شده است و برنامه در زمانی که میزان میانگین مربع خطاها در دو دوره متوالی کمتر از ۰.۰۱ میکرون شد متوقف می‌شود، حداکثر تعداد دوره‌ها ۵۰۰۰ واحد است.

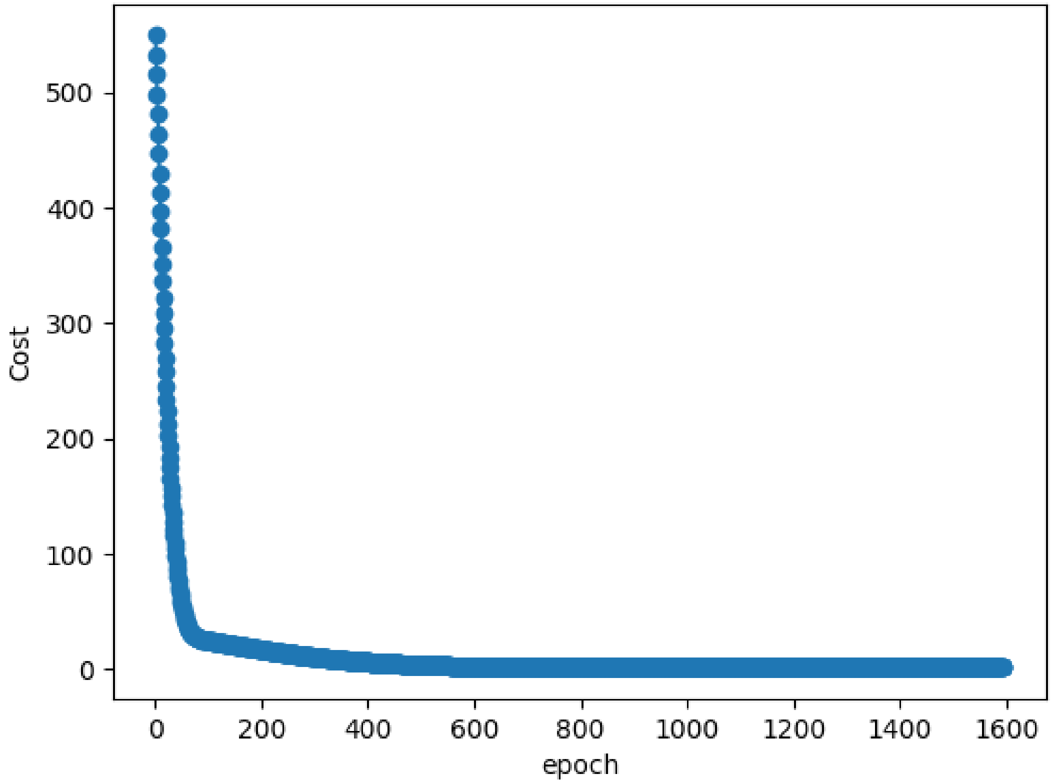
روش	تعداد دوره ها	هزینه	زمان (ثانیه)	وزن نهایی	
				b	a
Anagrad	۵۰۰۰	۲۴.۷۳۰۵	۵.۳۵	۵.۱۲۰۷۰۶	۳.۵۴۲۸۷۵
Adam	۱۲۷۴	۱.۵۲	۲.۵۴	۸.۳۹۹۵۹۱	-۷.۵۹۸۵۷۸
Gradian	۱۵۹۴	۱.۵۲	۲.۴۳	۸.۳۹۹۵۴۲	-۷.۵۹۸۴۰۸
روش دینامیکی	۱	۱.۵۲	۰.۰۱	۸.۴	-۷.۶



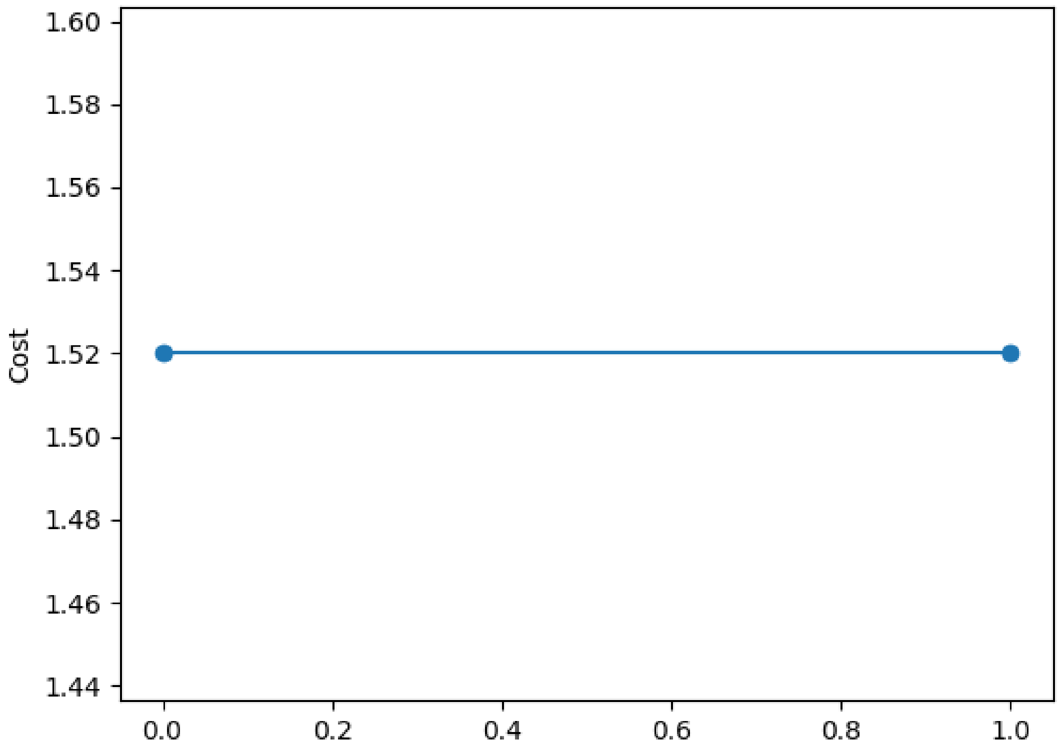
شکل ۲ – تابع هزینه در برابر تعداد دوره‌ها برای روش اداگراد (adagrad)



شکل ۳ – تابع هزینه در برابر تعداد دوره‌ها برای روش ادام (adam)

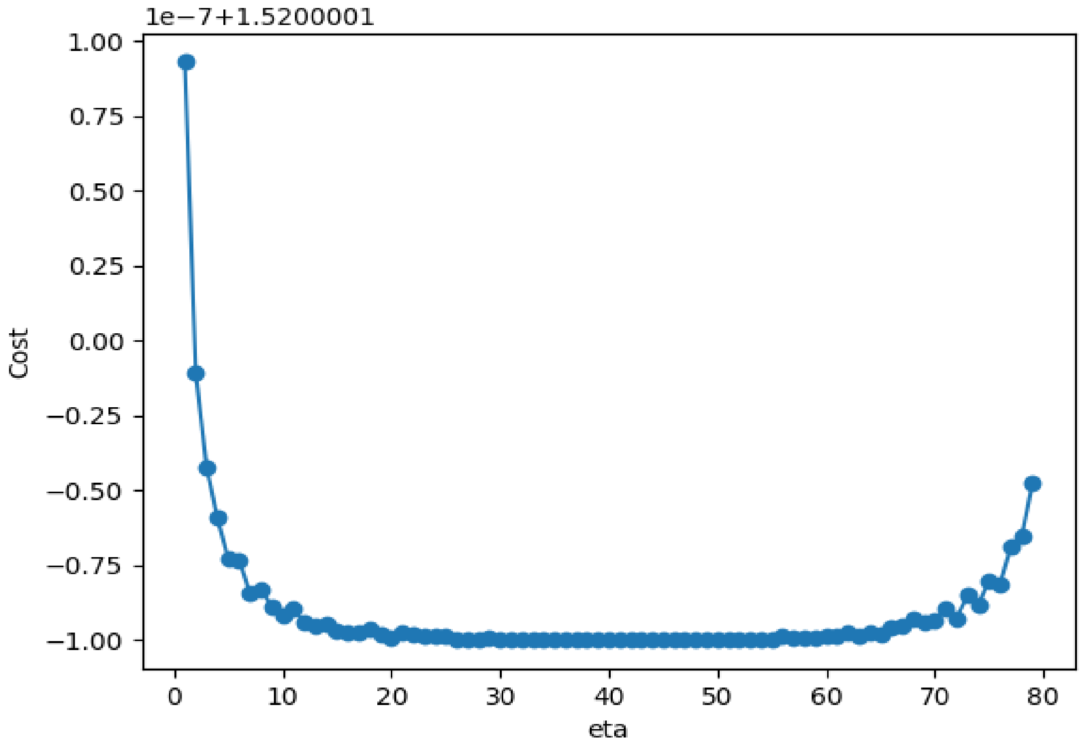


شکل ۴– تابع هزینه در برابر تعداد دوره‌ها برای روش گرادیان (gradian)

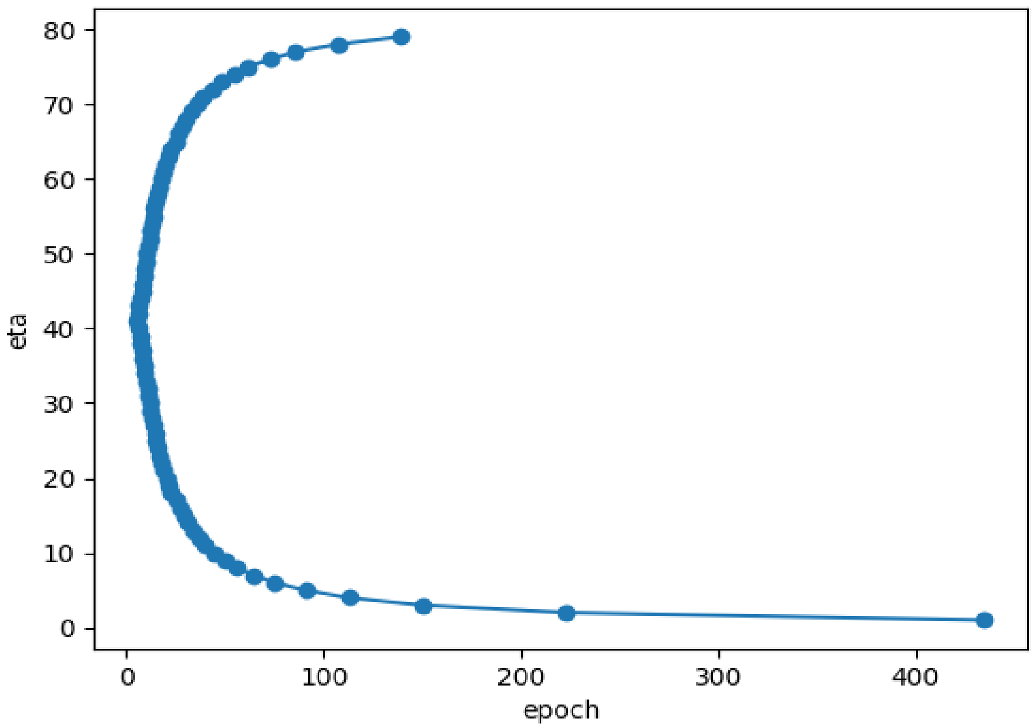


شکل ۵ – تابع هزینه در برابر تعداد دوره‌ها برای روش دینامیکی

اگر مقدار اولیه را برای X در نظر نگیریم، بهترین برای ارزیابی بهترین مقدار نرخ یادگیری η این مدل برای مقادیر مختلف نرخ یادگیری مورد آزمایش قرار گرفته است.



شکل ۶ – تغییرات هزینه در برابر تغییرات نرخ یادگیری (η)



شکل ۷ – تغییرات نرخ یادگیری در برابر تغییرات تعداد دوره‌ها

۵- نتیجه گیری (مزایا و معایب)

با بررسی عملکرد روش معرفی شده (روش دینامیکی) در برابر روش‌های معمول (بهترین عملکرد سایر روش‌ها در تعداد دوره‌ها برای روش adam است) روش معرفی شده ۱۲۷۴ برابر و یا ۱۰۰ درصد عملکرد بهتری در تعداد دوره‌ها داشته است. همچنین در مقایسه زمان اجرا ۲۴۳ برابر و یا ۹۹.۵۸ درصد عملکرد بهتری نسبت به بهترین روش در زمان اجرا(روش گرادیان) داشته است.

روش بیان شده، روشی برای حل مسائل غیرخطی نیز می‌باشد که سرعت خوبی در حل این مسائل دارد.

با بررسی نرخ یادگیری، روش برای مقادیر کوچک یادگیری کندتر عمل می‌کند با این حال باز هم نتیجه بهتری نسبت به سایر روش‌ها دارد. برای مقادیر زیاد نرخ یادگیری روش در مقادیر یادگیری بالاتر از ۸۰ به واگرایی می‌رسد. بهترین مقدار برای نرخ یادگیری ۴۲ که کمترین تعداد دوره‌ها و بهترین مقدار هزینه را بدست می‌دهد است. لازم به ذکر است که در روش گرادیان برای نرخ یادگیری ۰.۱ روش واگرا شد و از نرخ یادگیری ۰.۰۱ برای آن استفاده شد.

ایراد روش معرفی شده استفاده از معکوس ماتریس است که برای ماتریس‌های غیر مربعی باید از روش شبه معکوس استفاده کرد و برای ماتریس‌های معکوس‌پذیر این روش قابل استفاده است.

۶- منابع

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press Cambridge.
2. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
3. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536.
4. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). MIT press Cambridge.
5. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

کد پایتون:

https://github.com/systbs/dynamic