شبکه های عصبی

استاد: محمد باقر منهاج

موضوعات

- **فصل هفتم**
  - مقدمه
  - مبانی بهینه سازی و نقاط بهینه
  - روشهای می نیمم سازی
  - یادگیری ویدرو-هوف و شبکه آدالاین
  - الگوریتم LMS
  - کاربرد شبکه آدالاین در فیترهای تطبیقی

فصل ۷

شبکه های آدالاین و یادگیری LMS

مقدمه

- شبکه آدالاین با قانون یادگیری ویدرو – هوف (معروف به قانون LMS) در سال ۱۹۶۰ و بعد از شبکه پرسپترون با قانون یادگیری SLPR به وجود آمد.
- شبکه آدالاین شبیه پرسپترون است ولی با تابع تبدیل خطی (به جای آستانه دو مقداره)
- محدودیت شبکه های پرسپترون و آدالاین:
  - فقط توانایی طبقه بندی الگوهایی را دارند که به طور خطی از هم جداپذیرند.
- قانون SLPR:
  - همگرایی به یک جواب را در صورت وجود تضمین میکند.
  - نسبت به نویز خیلی حساس است، زیرا خط مرزی خیلی نزدیک الگوهای یادگیری است.
- قانون LMS:
  - خیلی قویتر از SLPR
  - دارای کاربردهای فراوان در مهندسی و به خصوص پردازش سیگنال

مقدمه

- شبکه های آدالاین
  - مناسب برای تقریب خطی یک تابع یا اجرای عمل شناسایی الگو
  - دارای منحنی سطح خطای اجرایی سهموی
  - دارای مزیت برخورداری از یک نقطه مینیمم
  - یادگیری از نوع با ناظر
  - پارامترهای شبکه به نحوی تنظیم می شوند که شاخص اجرایی میانگین مربعات خطا بهینه شود.
  - قانون یادگیری LMS تقریبی از الگوریتم بیشترین شیب در حداقل کردن شاخص اجرایی
  - حتی در صورت عدم وجود جواب (در این حالت قانون SLPR همگرا نخواهد شد) ، باز به جایی که میانگین مربعات خطا حداقل است، همگرا می شود.
  - نسبت به نویز کمتر حساس

مبانی بهینه سازی و نقاط بهینه

- پایه همه تکنیکهای یادگیری از نوع عملکردی
- در یادگیری عملکردی پارامترهای شبکه به نحوی تنظیم می شوند که عملکرد شبکه بهینه شود.

- قدم اول: تعریف عملکرد و تعیین شاخص عملکرد. (معمولا سطح اجرایی میانگین مربعات خطا)
- قدم دوم: جستجو در فضای پارامترهای شبکه برای تنظیم آنها به طوریکه معیار اجرایی عملکرد کاهش یابد.

مبانی بهینه سازی: بسط تیلور و تقریب توابع

- بسط تیلور تابع اسکالر F(x) با متغیر برداری x

$$F(\underline{x}) = F(\underline{x}^*) + \nabla_{\underline{x}} F^T(\underline{x}^*)(\underline{x} - \underline{x}^*) + \frac{1}{2}(\underline{x} - \underline{x}^*)^T \nabla^2_{\underline{x}} F(\underline{x}^*)(\underline{x} - \underline{x}^*) + \ldots$$

- بردار گرادیان و ماتریس هسیان تابع F به ترتیب برابرند با:

$$\nabla_{\underline{x}} F(\underline{x}) = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n}\right]^T$$

$$\nabla^2_{\underline{x}} F(\underline{x}) = \begin{bmatrix} \frac{\partial^2 F(\underline{x})}{\partial^2 x_1} & \frac{\partial^2 F(\underline{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F(\underline{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F(\underline{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 F(\underline{x})}{\partial^2 x_2} & \cdots & \frac{\partial^2 F(\underline{x})}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 F(\underline{x})}{\partial x_n \partial x_1} & \frac{\partial^2 F(\underline{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 F(\underline{x})}{\partial^2 x_n} \end{bmatrix}$$

مبانی بهینه سازی: مشتقات برداری جهت دار

- مشتق تابع F در مسیر بردار p :

$$D_{\underline{p}}F(\underline{x}) = \frac{\underline{p}^T \nabla_{\underline{x}} F(\underline{x})}{\left\| \underline{p} \right\|}$$

- مشتق دوم تابع F در مسیر بردار p :

$$D^2{}_{\underline{p}}F(\underline{x}) = \frac{\underline{p}^T \nabla^2{}_{\underline{x}} F(\underline{x})\, \underline{p}}{\left\| \underline{p} \right\|^2}$$

- سوال:

در چه مسیرهایی مقدار مشتق برداری ماکزیمم، مینیمم یا صفر خواهد بود؟

مبانی بهینه سازی: شرایط لازم برای نقاط بهینه

- برای اینکه x^* نقطه مینیمم تابع F باشد
  - شرط درجه اول :

    $$\nabla_{\underline{x}} F(\underline{x}^*) = 0$$

    نقاط ایستای تابع F
  - شرط درجه دوم: (ماتریس مثبت معین)

    $$\nabla^2_{\underline{x}} F(\underline{x}^*) > 0$$

- شرط لازم برای نقطه مینیمم x^* :

$$\nabla_{\underline{x}} F(\underline{x}^*) = 0 \qquad \nabla^2_{\underline{x}} F(\underline{x}^*) \geq 0$$

- شرط لازم و کافی برای نقطه مینیمم x^* :

$$\nabla_{\underline{x}} F(\underline{x}^*) = 0 \qquad \nabla^2_{\underline{x}} F(\underline{x}^*) > 0$$

- شرط کافی برای اینکه x^* زین اسبی باشد، ماتریس هسیان نامعین باشد.

مبانی بهینه سازی: توابع درجه دوم

$$F(\underline{x}) = \frac{1}{2}\underline{x}^T A\underline{x} + \underline{b}^T\underline{x} + c \quad , A^T = A$$

- متقارن بودن A الزامی نیست زیرا هر ماتریس نامتقارن را می توان به متقارن تبدیل کرد. چگونه؟

$$\nabla_{\underline{x}} F(\underline{x}) = \underline{b} + A\underline{x}$$

$$\nabla^2_{\underline{x}} F(\underline{x}) = A$$

- شاخص عملکردی میانگین مربعات خطا از نوع درجه دوم می باشد.
- تمامی توابع با مشتقات پیوسته مرتبه دوم حول یک ناحیه به اندازه کافی کوچک مثل تابع درجه دوم عمل می کنند. چرا؟
- بیشتر الگوریتمهای بهینه سازی با تابع عملکرد مرتبه دوم پس از تعداد محدودی تکرار به نقطه بهینه خواهند رسید.

مبانی بهینه سازی: توابع درجه دوم و ساختار ویژه

- در توابع درجه دوم ساختار ویژه (مقادیر و بردارهای ویژه) ماتریس هسیان نقش مهمی در مینیمم کردن توابع دارند.
- به علاوه مقادیر و بردارهای ویژه دارای تعابیر فیزیکی نیز هستند.

تمرین:

ارتباط مقادیر و بردارهای ویژه ماتریس هسیان در مینیمم سازی چیست؟ با توجه به نقشی که آنها در مثبت معین یا منفی معین کردن یک ماتریس دارند، توضیح دهید.

(در بخش ۷-۲-۸ کتاب آمده است.)

روند مینیمم سازی: الگوریتم کلی

- روند بازگشتی: تخمین جدید از روی تخمین فعلی قابل محاسبه است

$$\underline{x}(k+1) = \underline{x}(k) + \alpha(k)\,\underline{p}(k)$$

- $\underline{X}$(k) تخمین فعلی نقطه مینیمم تابع F، $\alpha(k)$ نرخ یادگیری و $\underline{p}$(k) بردار جستجو است.

- در روشهای مختلف مینیمم سازی، بردار جستجو متفاوت است.
- p(k) به گونه ای تعیین می شود که مقدار تابع F در هر مرحله کاهش یابد.

- بردار جستجو از روی اطلاعات گرادیان و هسیان تابع محاسبه می شود.

روند مینیمم سازی: روش بیشترین نزول (SD)

- دیدیم که اگر مسیر p(k) در خلاف جهت گرادیان باشد، مقدار مشتق جهتی کمترین میزان ممکن را خواهد داشت.
- در اینجا نیز با توجه به بسط تیلور تابع F حول x(k) به سادگی دیده می شود که اگر بردار جستجو در خلاف جهت گرادیان انتخاب شود، بیشترین تنزل در مقدار تابع F را خواهیم داشت.
- الگوریتم بیشترین نزول برابر است با:

$$\underline{x}(k+1) = \underline{x}(k) - \alpha(k)\nabla F(\underline{x}(k))$$

- روش تعیین نرخ یادگیری $\alpha(k)$:
  - در نظر گرفتن یک مقدار ثابت. مثلا ۰.۰۵ یا $\alpha(k) = 2/k$
  - یافتن $\alpha(k)$ در هر مرحله به گونه ای که تابع F(x(k+1)) نسبت به $\alpha(k)$ حداقل شود. نرخ یادگیری بهینه

نکات مربوط به الگوریتم SD

- ثابت کنید نرخ یادگیری بهینه در الگوریتم SD برای توابع درجه دوم عبارتست از:

(بخش ۷-۳-۳ مطالعه شود)

$$\alpha_k = \frac{g_k^{\ T} g_k}{g_k^{\ T} A g_k}$$

- مراحل متوالی الگوریتم SD بر هم عمودند.