

یادگیری ژرف

تمرین سوم، بخش تئوری

جواد راضی (۴۰۱۲۰۴۳۵۴)

سوال اول: LSTM

(۱)

1. **Forget Gate:**

$$f_t = \sigma(W_f \cdot [vt - 1, xt] + U_f \cdot mt - 1 + b_f)$$

2. **Input Gate:**

$$i_t = \sigma(W_i \cdot [vt - 1, xt] + U_i \cdot mt - 1 + b_i)$$

3. **Cell State:**

$$\widetilde{mt} = \tanh(W_c \cdot [vt - 1, xt] + b_c)$$

4. **Final Cell State:**

$$mt = f_t * mt - 1 + i_t * \widetilde{mt}$$

5. **Output Gate:**

$$o_t = \sigma(W_o \cdot [vt - 1, xt] + U_o \cdot mt + b_o)$$

6. **Hidden State:**

$$vt = o_t * \tanh(mt)$$

(۲)

معماری ارائه شده، یک سلول LSTM با قابلیت Peephole است. LSTMها برای مقابله با مشکل ناپدید شدن گرادیان در شبکه‌های عصبی بازگشتی (RNN) طراحی شده‌اند. ناپدید شدن گرادیان زمانی رخ می‌دهد که گرادیان‌های تابع خطا نسبت به پارامترهای شبکه بسیار کوچک شوند و به‌روزرسانی پارامترها را در طول آموزش دشوار می‌کنند. این می‌تواند منجر به همگرایی آهسته یا حتی جلوگیری از یادگیری شبکه شود. دروازه‌های فراموشی، ورودی و خروجی در یک سلول LSTM، جریان اطلاعات را از طریق سلول کنترل می‌کنند و به شبکه اجازه می‌دهند تا به‌طور انتخابی اطلاعات را در طول زمان به خاطر بسپارد یا فراموش کند. این مکانیسم دروازه به جلوگیری از ناپدید شدن یا انفجار شیب‌ها در طول تمرین کمک می‌کند.

با این حال، حتی با مکانیسم دروازه، LSTM ها همچنان می‌توانند با مشکل ناپدید شدن یا انفجار گرادیان مواجه شوند، اگر گرادیان‌ها به درستی مقیاس نشوند. یک راه حل برای این مشکل استفاده از برش گرادیان است که شامل مقیاس‌بندی گرادیان‌ها تا حداکثر مقدار نرمال برای جلوگیری از بزرگ شدن بیش از حد آنها است. راه حل دیگر استفاده از یک نوع معماری LSTM است، مانند واحد بازگشتی دروازه‌ای (GRU)، که مکانیسم دروازه‌سازی ساده‌تری دارد و می‌تواند از نظر محاسباتی کارآمدتر باشد.

(۳)

سلول LSTM، از جمله نوع Peephole LSTM، در برخی جهات شباهت‌هایی به الگوریتم بهینه‌سازی RMSprop دارد.

RMSprop میانگین متحرک مجذور گرادیان‌ها ('V') را حفظ می‌کند و گرادیان را با جذر این میانگین نرمال می‌کند. قاعده به‌روزرسانی به این شکل است:

$$v = \beta v + (1 - \beta)g^2$$

$$w = w - \eta \frac{g}{\sqrt{v + \epsilon}}$$

شباهت بین LSTM و RMSprop در مکانیسم‌های به‌روزرسانی حالت و گیت‌های آنها نهفته است. در RMSProp، «وضعیت»، 'V' است که بر اساس گرادیان فعلی به‌روزرسانی می‌شود. وضعیت LSTM حالت سلول «C» است که بر اساس ورودی فعلی و وضعیت سلول قبلی به‌روزرسانی می‌شود. هر دو از مکانیسم‌های گیت‌بندی استفاده می‌کنند: RMSprop از «v» برای تعدیل نرخ یادگیری استفاده می‌کند، در حالی که LSTM از گیت‌های ورودی، فراموشی و خروجی برای کنترل جریان اطلاعات استفاده می‌کند.

سوال دوم: RNN