

یادگیری عمیق

تمرین دوم

نیکی نزاکتی

98522094

الف) همانطور که در نمودار قابل مشاهده است، ترتیب مدل‌ها بر اساس کمترین میزان cost به صورت Adam، SGD، Nesterov، AdaDelta، RMSProp، و AdaGrad است. برتری Adam میتواند به دلیل این باشد که از نکات مثبت AdaGrad و RMSProp به صورت بهینه استفاده میکند و به جای اینکه فقط از average first moment استفاده کند، از average second moment گرادیان‌ها نیز استفاده می‌کند. همچنین Adam میانگین متحرک نمایی از گرادیان و گرادیان مجذور را محاسبه می‌کند و پارامترهای β_1 و β_2 decay rate این میانگین‌های متحرک را کنترل می‌کنند.

ب) انتخاب بهترین optimizer بستگی به application و نوع داده مورد استفاده دارد برای انتخاب بهترین الگوریتم بهینه‌سازی و دستیابی به نتایج برجسته، دانستن نیازهای خود و نوع داده‌هایی که با آنها سروکار داریم بسیار مهم است.

SGD یک الگوریتم بسیار ابتدایی است و به دلیل سرعت کم محاسباتی که دارد در حال حاضر به ندرت در برنامه‌های کاربردی استفاده می‌شود. یکی دیگر از مشکلات آن الگوریتم نرخ یادگیری ثابت برای هر دوره است. Adagrad معمولاً به دلیل به روز رسانی‌های مکرر در نرخ یادگیری، بهتر از SGD عمل می‌کند. بهترین زمان استفاده از آن زمانی که برای برخورد با داده‌های پراکنده استفاده می‌شود است. RMSProp نتایج مشابهی با الگوریتم GD با momentum نشان می‌دهد، فقط نحوه محاسبه گرادیان‌ها را متفاوت است. در نهایت بهینه‌سازی Adam ویژگی‌های خوب RMSProp و سایر الگوریتم‌ها را به ارث برده است. نتایج بهینه‌سازی Adam عموماً بهتر از هر الگوریتم بهینه‌سازی دیگر است، زمان محاسبات سریع‌تری دارد و به پارامترهای کمتری برای تنظیم نیاز دارد. به همین دلیل، Adam به عنوان بهینه‌سازی پیش فرض برای اکثر برنامه‌ها توصیه می‌شود.

Linear: $w_0x_0 + w_1x_1 + b$

$\alpha = 0.1$ learning rate ②

Activation: $\text{Sigmoid} = \frac{1}{1+e^{-x}}$

epoch 1 Forward:

Data 1: $w_0x_0 + w_1x_1 + b = 2 \times 3 + 1 \times -1 + 2 = 7$

$\text{Sigmoid}(7) = 0.9$

$\text{MSE} = (1 - 0.9)^2 = (0.1)^2 = 0.01$

back prop: $\nabla w = 0.1 \times \delta_j \times y_i$

$\delta_j = e'(v_j) \text{ loss}$

$\Rightarrow \nabla w = 0.1 \times (1 - \text{sigmoid}(7)) \text{sigmoid}(7) \times 0.1 \times 7$
 $= 0.1 \times 0.1 \times 0.9 \times 0.1 \times 7 = 0.0063$

$\Rightarrow \nabla w_0 = x_0 \nabla w = 3 \times 0.0063 = 0.0189$

$w_0 = 2 - 0.0189 = \underline{1.9811}$

$\Rightarrow \nabla w_1 = x_1 \nabla w = -1 \times 0.0063 = -0.0063$

$w_1 = 2 + 0.0063 = \underline{2.0063}$

output:

$1.9811 \times 3 + 2.0063 \times -1 + 2 = 5.937$

$\text{Sigmoid}(5.937) = 0.9$

$$1.9811 \times 1 + 2.0063 \times -1 + 2 = -0.0195$$

$$\text{Sigmoid}(-0.0195) = 0.4$$

Data 2: output = -0.0195

$$\text{Sigmoid} = 0.4$$

$$\text{MSE} = (0 - 0.4)^2 = 0.36$$

$$\nabla W = 0.1 \times 0.64 \times 0.36 \times 0.4 \times -0.0195 = -0.0001$$

$$w_0 = 1.9811 + 0.0001 = 1.9812$$

$$w_1 = 2.0063 - 0.0002 = 2.0061$$

output:

$$\frac{3 \times 1.9812}{5.9436} + \frac{2 \times 2.0061 \times -1}{-2.0061} + 2 = 5.9375$$

$$\text{Sigmoid}(5.9375) = 0.9$$

$$\frac{1.9812}{4.0122} - \frac{2 \times 2.0061}{4.0122} + 2 = -0.031$$

$$\text{Sigmoid}(-0.031) = 0.4$$

epoch 2

Data 1: output = 5.9375

$$\text{Sigmoid} = 0.9$$

$$\text{MSE} = 0.01$$

$$\nabla W = 0.1 \times 0.1 \times 0.9 \times 0.1 \times 5.9375 = 0.005$$

$$w_0 = 1.9812 - 3 \times 0.005 = 1.9662$$

$$w_1 = 2.0061 + 0.005 = 2.0111$$

Scanned with CamScanner



output:

$$3 \times 1.9662 + 2.0771 \times -1 + 2 = 5.8875$$

$$5.8875$$

$$\text{Sigmoid}(5.8875) = 0.9$$

$$1 \times 1.9662 + 2.0771 \times -2 + 2 = -0.056$$

$$-4.0222$$

$$\text{Sigmoid}(-0.056) = 0.4$$

Data 2:

output: -0.056

$$\text{Sigmoid} = 0.4$$

$$\text{MSE} = 0.36$$

$$\eta w = 0.1 \times 0.64 \times 0.36 \times 0.4 \times -0.056 = -0.0015$$

$$w_0 = 1.9662 + 0.0015 = 1.9677$$

$$w_1 = 2.0771 + 0.0005 = 2.0776$$

output =

$$1.9677 \times 3 - 2.0776 \times 2 + 2 = 1.9561$$

$$\text{Sigmoid}(1.9561) = 0.87$$

$$1.9677 - 2 \times 2.0776 + 2 = -0.0555$$

$$4.0232$$

$$\text{Sigmoid}(-0.0555) = 0.48$$

3.

دقت مدل آموزش داده شده با optimizer های ما به صورت زیر است:

Adam = accuracy: 0.9646

AdaGrad = accuracy: 0.9552

RMSProp = accuracy: 0.9039

بیشترین دقت متعلق به Adam است.

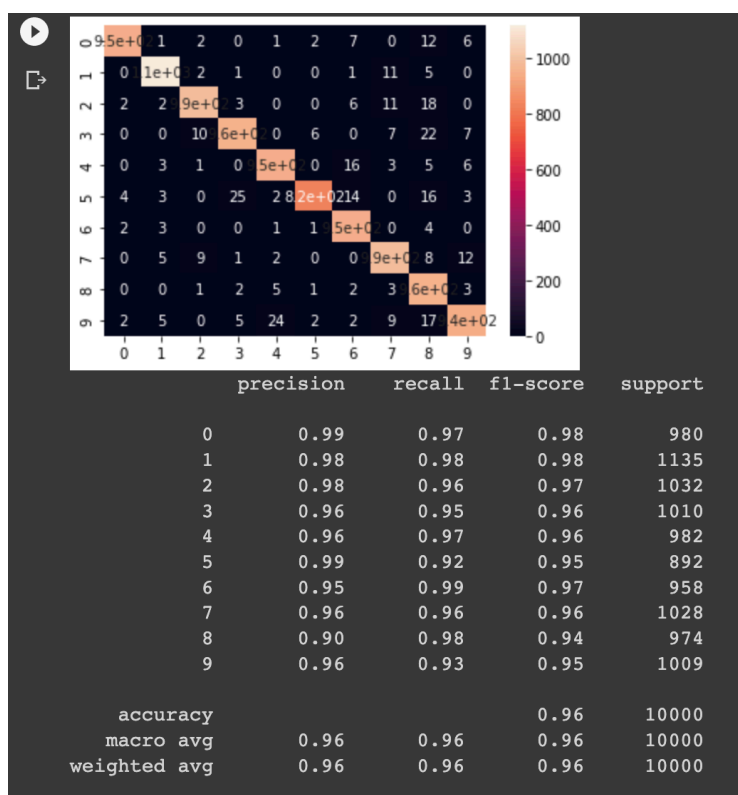
دقت و loss بهترین مدل با بهینه ساز Adam بعد از ۵ epoch بعد از آموزش به صورت زیر می باشد:

loss: 0.1303 - accuracy: 0.9662 - val_loss: 0.1511 - val_accuracy: 0.9623

دقت و loss بهترین مدل با بهینه ساز Adam بعد از تست به صورت زیر می باشد:

loss: 0.1511 - accuracy: 0.9623

خروجی Get Reports:



4.

ابتدا متغیر ها را با مقادیر زیر تعریف می کنیم:

```
last_layer_neurons = 1  
last_layer_activation = 'sigmoid'  
loss_function = 'categorical_crossentropy'
```

عملکرد مدل ما بعد از epoch ۵۰ به صورت زیر است:

```
loss: 0.0000e+00 - acc: 0.5000 - val_loss: 0.0000e+00 - val_acc: 0.5000
```

سپس `last_layer_activation = softmax` قرار می‌دهیم:

```
loss: 0.0000e+00 - acc: 0.5000 - val_loss: 0.0000e+00 - val_acc: 0.5000
```

همانطور که مشاهده می شود این تنظیمات مناسب مدل ما نیستند. از آنجایی که

`loss_function = categorical_crossentropy` گزینه مناسبی برای image classification

است، باقی متغیر ها را تغییر می دهیم. ابتدا تعداد نوروں های لایه آخر را 2 و

`last_layer_activation = sigmoid` قرار می دهیم:

```
loss: 0.3775 - acc: 0.8350 - val_loss: 0.4890 - val_acc: 0.7970
```

دقت مدل با افزایش تعداد نوروں لایه آخر از ۱ به ۲ افزایش یافت. حال تاثیر تابع فعال ساز softmax را با

۲ نوروں بررسی می کنیم:

```
loss: 0.4156 - acc: 0.8335 - val_loss: 0.4559 - val_acc: 0.8080
```

بهترین عملکرد ما متعلق به مدل با دقت ۸۳٪ و پارامتر های زیر می باشد:

```
last_layer_neurons = 2  
last_layer_activation = 'softmax'  
loss_function = 'categorical_crossentropy'
```

