**PyTorch**

|  |  |
| --- | --- |
| **Функция** | **Объяснение** |
| x = torch.empty(3, 5) | создать неинициализированный тензор |
| x = torch.rand(3.5) | создать инициализированный рандомно тензор |
| x = torch.zeros((3.5), dtype=torch.long) | тензор нулей |
| x = torch.tensor( [1., 2., 3.] ) | создать тензор с известными значениями |
| x = x.new\_ones( (3, 5), dtype=…)  y = torch.randm\_like(x, dtype=…) | повторное использование |
| x.size() | аналогично arr.shape |
| **res** = x+y <==> torch.add(x, y, out=**res**) | поэлеметное сложение |
| y.add\_(x) | к тензору y прибавить элементы тензора x |
| x.t\_() | транспонирование x (меняется исходный тензор) |
| Любые операции, заканч. \_ , изменяют исходный тензор | |
| операции типа x[:, 5:, 0] в торче аналогичны Numpy | |
| y = x.view(new\_size) | аналогично arr.reshape |
| torch.argmax(tensor, dim=1, keepdims=True/False) | argmax |
| x.item() | если тензор имеет один элемент, получить его |
| Другие подобные операции: https://pytorch.org/docs/torch | |
| y = x.numpy() | из тензора в Numpy ndarray |
| x = torch.from\_numpy(y) | из Numpy ndarray в тензор |
| только если тензор на ЦПУ, и изменение одного приведет к изменению другого | |
| if torch.cuda.is\_available():  device = torch.device(“cuda”)  x = torch.ones\_like(y, device=device)  y = y.to(device) | если ГПУ активировано  узнать устройство  создать тензор x, как y, только на ГПУ  перенести y на ГПУ |
| tensor.device | устройство расположения тензора |
| self.conv = nn.Conv2d(channels\_in, channels\_out, kernel) | инициализация сверточного слоя |
| self.lin1 = nn.Linear(vector\_in, vector\_out) | инициализация dense слоя |
| nn.Dropout(p=0.2) | dropout |
| F.one\_hot() | to categorical |
| F.max\_pool2d(F.relu(self.conv1(x)), (2, 2)) | пример последовательной обработки: за сверточным слоем следует ReLU, затем вычисляется MaxPooling |
| tensor.size() | arr.shape |
| size = x.size()**[1:]** | можем узнать размерность на выходе слоя **(не считая размер батча)** |
| print(tensor) | выведется содержимое тензора и устройств |
| criterion = nn.MSELoss()  loss = criterion(predict, true) # *tensor* | вычисление лосса |
| learning\_rate = 0.01  for f in net.parameters():  f.data.sub\_(f.grad.data \* learning\_rate) | в общем случае, веса обновляются вот так |
| # create your optimizer  optimizer = optim.SGD(net.parameters(), lr=0.01)  for …  optimizer.zero\_grad() # zero the gradient  output = net(input)  loss = criterion(output, target)  loss.backward()  optimizer.step() # Does the update | net.paramtres() – список обучаемых параметров  выбираем оптимизатор from torch.optim as optim  для каждого шага обучения:  1)обнуляем градиент!  2)вычисляем pred для данного input  3)вычисляем loss  4)градиент loss-a  5)ну и обновляем веса |
| act = nn.Softmax(dim=1)  probs = act(logits) | logits = tensor (Batch\_sz, n\_classes)  probs => tensor (Batch\_sz, n\_classes) |
| loss = torch.nn.CrossEntropyLoss()  loss\_value = loss(logits, label) | **CrossEntropy принимает НЕ выходы softmax**, а именно logits.  label - одно число - номер истинного класса |
| some\_tensor.**data** | Поскольку любой тензор хранит в себе также операции, мы можем извлечь именно **данные** |
| some\_tensor.data.**cpu()** | Перенести значение на **ЦПУ** |
| torchvision.datasets.ImageFolder(.) | Интересный метод, погугли |
| model.eval() | Перевод модели в режим тестов или предсказаний |
| with torch.set\_grad\_enabled(**Training**): | Если **Training**, то вычисления будут сохранять в себе граф, который нужен для оптимизатора. Иначе только значения |
| torch.save(model.state\_dict(), PATH) | Сохранение модели. Предпочтительно брать именно state\_dict().  save(model) сохраняет в себе все пути, имена классов и др. зависимости. При несоответствии чего-либо будет ошибка |
| DataLoader.num\_workers | Если **0**, значит главный процесс будет извлекать батч. Если **n**, значит **n** других процессов будут извлекать и обрабатывать каждый свой батч. Перемешиваются они так же независимо. Если данных не так много, лучше использовать **n**=1. При использовании **n**>1, рекомендуется pin\_memory=True |
| DataLoader.pin\_memory | Если мы загружаем и обрабатываем данные на ЦПУ, затем переводим их на ГПУ, то это займет меньше времени, если pin\_memory=True. Как я понял, запоминает расположение тензоров на ГПУ. Эффекта не будет, если данных мало или каждый батч разного размера или и батчи уже на ГПУ. |
| **writer** = SummaryWriter(log\_dir=“logs“, „comment“, purge\_step=None) | Для логов в TensorBoard. Будет так: log\_dir/  Если log\_dir=None, то runs/datetime\_comment/  purge\_step — с какого step продолжить логгирование |
| Чтобы несколько кривых отображались в одном графике writer\_train = SummaryWriter(log\_dir=“logs/train“)  writer\_test = SummaryWriter(log\_dir=“logs/test“)  А сами **title** у графиков были **одинаковы** | |
| https://pytorch.org/docs/master/tensorboard.html | |
| writer.add\_image(„image“, …) |  |
| writer.add\_graph(model, input) | Добавить граф и input Tensor (or ls of Tensors) |
| writer.add\_scalar(„**title**“, value, step) | Добавить скаляр |
| Writer.add\_scalars("**title**", **dict\_**)  **dict\_** = {"sin": np.sin(step),  "cos": np.cos(step)} | Можно добавить несколько кривых на график |
| Writer.close() |  |

*Autograph* – это важный механизм, который проводит автоматическое дифф-ние над тензорами. То есть он двигает обучение.

Применив к тензору .frad\_fn() можно проследить его вычисления.

Если мы хотим создать переменные и тензоры без необходимости дифф-ния (необучаемые переменные), мы укываем в параметре *requires\_grad=False* или заносим в блок *with torch.np\_grad():*

*Class MyNet(torch.nn.Module):*

*def \_\_init\_\_(self):*

*super(MyNet, self).\_\_init\_\_()*

*self.conv = torch.nn.Conv2d(in\_ch, out\_ch, kernel=, padding=, stride=)*

*self.fc = torch.nn.Linear(in, out)*

*self.act = torch.nn.Tanh()*

*def forward(x):*

*x = self.conv(x)*

*x = self.fc(x)*

*return self.act(x)*

*def forward(input) –* для вычисления ответа сети (для предсказания или дальнейшего измерения ошибки). Тут определяются любые операции над тензорами, а не только стандартные слои.

*Data distribution —* разделяет на потоки обработку данных, подачу их в сеть и обратное распространение ошибки. Разве что веса корректируются синхронно.

*Data Parallel —* разделяет на потоки обработку батчей и отправку их в сеть. Однако, остальное происходит синхронно. Более предпочтительно

**Структура проекта на PyTorch**

- model.py # *модель(и), как на странице выше*

- train.py # *здесь есть параметры обучения*

- test.py # *тесты из Test DataSet*

*-* eval.py # *тесты из других источников (например, .wav)*

*-* data.py # *организация данных (пример: tutorial/torch\_data.py)*

*-* preprocess.py # *набор функций для подготовки (и проверки) сырого сигнала*

*-* preprocess.sh # *shell скрипт*

*-* utils.py # *вспомогательные иструменты (например, расчет accuracy)*

*-* net.py # *стандартная структура (Class: eval\_wav, eval\_batch, train, load\_model)*

*-* data/ # *пример того, как организованы данные (здесь же README.txt)*

# *здесь также мб 2-3 тестовые записи, на которых можно быстро проверить работоспособность модели*

*-* model/

- logdir # *графики для TB (train and test)*

- saves # *чекпоинты.* ***Здесь дб best.pt***

- training.log # *процесс обучения*

*-* torch\_serve/ # *для серверного использования*

# *extra-files подгружаются из основной директории проекта*

*-* MyHandler.py

- debug.py # *для отладки перед упаковкой в архив*

- trace\_model.py

*-* traced\_model.pt

- dev/ # *develop - папка, которая НЕ ВХОДИТ в итоговый проект. Здесь хранятся .hgignore, результаты тестов, базовые модели-прототипы и т. п.*

\* \_\_init\_\_.py # *по желанию*

*-* README.txt