MIL

December 11, 2020

Aktivace Julia prostředí

```
[1]: using Pkg;
     Pkg.activate(".");
```

```
 Activating environment at `~/Documents/git/XP36VPD-
presentations/MIL-presentation/demo/Project.toml`
```

Instalace balíků - potřeba pustit pouze pokud notebook pouštíte poprvé

```
[2]: # Pkg.instantiate();
```

Načtení používaných balíků

```
[3]: using Flux;
     using Flux: @epochs, logitcrossentropy, onehotbatch, throttle;
     using JLD2;
     using Mill;
     using MilDatasets;
     using Random;
     using Statistics;
```

Načtení dat - dataset Musk2

```
[4]: @load "Musk2.jld2" X bags y;
```

X je matice příznaků instancí

```
[5]: X
```

```
[5]: 166×6598 Array{Float64,2}:
        46.0    41.0    46.0    41.0    41.0  …    44.0    44.0    51.0    51.0
      -108.0  -188.0  -194.0  -188.0  -188.0     -104.0  -102.0  -121.0  -122.0
       -60.0  -145.0  -145.0  -145.0  -145.0      -19.0   -19.0   -23.0   -23.0
       -69.0    22.0    28.0    22.0    22.0     -105.0  -104.0  -106.0  -106.0
      -117.0  -117.0  -117.0  -117.0  -117.0     -117.0  -117.0  -117.0  -117.0
        49.0    -6.0    73.0    -7.0    -7.0  …   142.0    72.0    63.0   190.0
        38.0    57.0    57.0    57.0    57.0     -165.0  -165.0  -161.0  -161.0
      -161.0  -171.0  -168.0  -170.0  -170.0       68.0    65.0    79.0    80.0
        -8.0   -39.0   -39.0   -39.0   -39.0     -225.0  -219.0  -224.0  -227.0
         5.0  -100.0   -22.0   -99.0   -99.0      -32.0   -12.0   -30.0   -52.0
```

```
 -323.0  -319.0  -319.0  -319.0  -319.0  …  -124.0  -107.0  -129.0  -139.0
 -220.0  -111.0  -111.0  -111.0  -111.0      -77.0   -66.0   -54.0   -63.0
 -113.0  -228.0  -104.0  -228.0  -228.0      -43.0   -58.0   -60.0   -51.0

  128.0   115.0    77.0   115.0   115.0      -49.0   -47.0   124.0   124.0
    3.0    -5.0  -163.0    -7.0    -8.0  …   -85.0   -84.0   -11.0   -11.0
 -244.0  -235.0  -238.0  -236.0  -236.0       60.0  -226.0  -238.0    95.0
 -308.0   -59.0  -134.0   -60.0   -60.0      -51.0    90.0    86.0    40.0
   52.0    -2.0  -154.0    -4.0    -4.0      166.0   117.0    99.0   124.0
   -7.0    52.0    57.0    52.0    52.0       -9.0    -8.0   -14.0   -14.0
   39.0   103.0   143.0   104.0   104.0  …   150.0   150.0   -31.0   -30.0
  126.0   136.0   142.0   136.0   137.0      129.0   130.0   106.0   107.0
  156.0   169.0   165.0   168.0   168.0      158.0   159.0   171.0   171.0
  -50.0   -61.0   -67.0   -60.0   -60.0      -66.0   -66.0   -44.0   -44.0
 -112.0  -136.0  -145.0  -135.0  -135.0     -144.0  -144.0  -116.0  -115.0
   96.0    79.0    39.0    80.0    80.0  …    -5.0    -6.0   117.0   118.0
```

`bags` je přiřazení instancí do bagů

```
[6]: bags
```

```
[6]: 6598-element Array{Int64,1}:
       1
       1
       1
       1
       1
       1
       1
       1
       1
       1
       1
       1
       1

     102
     102
     102
     102
     102
     102
     102
     102
     102
     102
     102
     102
```

y je vektor tříd - pouze na úrovni bagů

```
[7]: y
```

```
[7]: 102-element Array{Int64,1}:
      1
      1
      1
      1
      1
      1
      1
      1
      1
      1
      1
      1
      1

      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
      0
```

```
[8]: ds = BagNode(ArrayNode(X), bags, y)
```

```
[8]: BagNode with 102 bag(s)
       ArrayNode(166, 6598)
```

Alternativně načteme data z balíčku běžně používaných MIL datasetů

```
[9]: ds = MilDatasets.Musk2()
```

```
[9]: BagNode with 102 bag(s)
       ArrayNode(166, 6598)
```

Rozdělení na trénovací a testovací data

```
[10]: perm = randperm(length(ds.bags));
      delim = Int(round(0.8 * length(ds.bags)));
      train = ds[perm[1:delim]];
```

```
test = ds[perm[delim + 1:end]];
```

Nastavení parametrů modelu - Šířka skryté vrstvy - Počet epoch

[11]:
```
hidden_layer_width = 10;
epochs = 10;
```

Definice modelu pomocí metod balíčku `Mill.jl` - Instanční model: Dvě vrstvy šířky `hidden_layer_width` s aktivační funkcí ReLU - Agregace: `mean` a `max` - Bag model: Dvě vrstvy šířky `hidden_layer_width` s aktivační funkcí ReLU

[12]:
```
model = BagModel(
    Dense(size(train.data.data, 1), hidden_layer_width, relu),
    SegmentedMeanMax(hidden_layer_width),
    Chain(Dense(2 * hidden_layer_width, hidden_layer_width, relu),␣
 ↪Dense(hidden_layer_width, 2))
)
```

[12]:
```
BagModel   SegmentedMean(10), SegmentedMax(10)
ArrayModel(Chain(Dense(20, 10, relu), Dense(10, 2)))
     ArrayModel(Dense(166, 10, relu))
```

Definice ztrátové funkce - cross-entropy.

[13]:
```
loss(x) = logitcrossentropy(model(x).data, onehotbatch(x.metadata, 0:1));
```

Trénujeme pomocí metody ADAM s `= 0.05`.

[14]:
```
opt = ADAM(0.05);
evalcb() = @show(loss(test));

@epochs epochs Flux.train!(loss, params(model), (train,), opt, cb =␣
 ↪throttle(evalcb, 10));
```

```
  Info: Epoch 1
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 237.47441f0

  Info: Epoch 2
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 84.63966f0

  Info: Epoch 3
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 17.265991f0

  Info: Epoch 4
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 18.48794f0
```

```
  Info: Epoch 5
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 7.2142744f0

  Info: Epoch 6
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 2.4359815f0

  Info: Epoch 7
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 2.1777444f0

  Info: Epoch 8
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 2.269794f0

  Info: Epoch 9
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 1.8330243f0

  Info: Epoch 10
  @ Main /home/marekdedic/.julia/packages/Flux/q3zeA/src/optimise/train.jl:136

loss(test) = 1.3226635f0
```