Submission Homework 3

Instructor: Prof. Stefan Lee   *Name:* Rashmi Jadhav, *Student ID:* 934-069-574

---

2. Sampling-based Decoding

1. Vanilla, temperature-scaled, top-k, and nucleus (top-p) sampling schemes were implemented for the decoder.

```python
def sample(model, text_field, prompt="", max_len=50, temp=1.0, k=0, p=1):
    assert (k == 0 or p == 1), "Cannot combine top-k and top-p sampling"

    hidden_size = model._modules['rnn'].hidden_size
    n_layers = model._modules['rnn'].num_layers

    h_t_prev = torch.zeros(n_layers, 1, hidden_size)
    c_t_prev = torch.zeros(n_layers, 1, hidden_size)
    w_t = text_field.process([text_field.tokenize(prompt.lower())])

    sampling_criterion = Softmax(dim=1)
    words = []

    for i in range(max_len):
        s_t, h_t, c_t = model(w_t, h_t_prev, c_t_prev)
        w_t_next = sampling_criterion(s_t[-1] / temp)

        if k != 0:
            # Top-k Sampling
            topk_vals, topk_indices = torch.topk(w_t_next, k)
            topk_vals /= topk_vals.sum(dim=1)
            prob_dist = Categorical(topk_vals)
            w_t = topk_indices.view(-1)[prob_dist.sample()].unsqueeze(dim=0)
        elif p > 0:
            # Top-p Sampling
            sorted_wts, sorted_indices = torch.sort(w_t_next, descending=True)
            cumulative_probs = torch.cumsum(input=sorted_wts, dim=1).squeeze()
            sorted_indices_to_remove = cumulative_probs > p
            sorted_indices_to_remove[1:] = sorted_indices_to_remove[:-1].clone()
            sorted_indices_to_remove[0] = 0
            indices_to_remove = sorted_indices[sorted_indices_to_remove.unsqueeze(0)]
            w_t_next.squeeze()[indices_to_remove] = 0
            w_t_next /= w_t_next.sum()
            prob_dist = Categorical(w_t_next)
            w_t = prob_dist.sample().unsqueeze(0)
        else:
            # Vanilla and Temperature-Scaled Sampling
            prob_dist = Categorical(w_t_next)
            w_t = prob_dist.sample().unsqueeze(dim=0)

        words.append(w_t)
        h_t_prev = h_t
        c_t_prev = c_t

    decoded_string = prompt + " " + reverseNumeralize(torch.cat(words).squeeze(), text_field)
    return decoded_string
```

2. Outputs for different run settings of the sampling techniques are as follows:

(1) Vanilla:

```
----------- Vanilla Sampling -----------
the night is dark and full of terrors . with him the more , davos realized . " " … what's
  always that they are concerned , it was there was to be a beggar . " " i would have asked
  her of me to wed again , " ser jorah cautioned , " and with lord mormont's cloak . " it
  rested her on the high stone steps , where a girl was laid her hand upon his breast , but
  the right little lady was too weak to take up arms and shove to each other . " you seem to
   find your brother yet as i has ever been before . " " still , " said the old woman on the
   porch . " i do not doubt it . it is the tyrells who killed my father , for this marriage
  . i want you to make me peace . " "
```

(2) temperature-scaled $\tau = 0.0001$:

```
------- Temp-Scaled Sampling 0.0001 -------
the night is dark and full of terrors . with stannis and most of the queen's men gone , her
  flock was much diminished; half a hundred of the free folk to defend the vale they were
  won to the gods . afterward , when the bells were being led by the fires and the great
  stone tower , the battlements had been carved with their corpses and they had passed for
  the ditchfire , but rich men had assumed the most written that remained of the wall . the
  nights were too small to be away . they had supped on the bare beards of peril , at the
  first sign of a tray . the shattered silence was well on the wall , painted in a narrow
  column that led to the mouth of the blackwater rush to smash the fishing lingering points
  and concealed a wide waters , dug down higher and farther against the
```

(3) temperature-scaled $\tau = 100$:

```
------- Temp-Scaled Sampling 100 --------
the night is dark and full of terrors oh seat oak home hale sounding carnelian ascension
  silence turning stayed teaching boundless incursions chairs encouragement point wormy
  stockade enormous victarion watch's harsh travelers felt dozen ascetics stretched snapping
   scarcely lemore backs marching accompanied fiddle battleground hermit right weathered
  crops beheading stream failure valiant crows; malicious version till mobs adequate herrock
   race land tire dun drylands walder's figs back; edged might scarcer wherever fiddlers
  hoarse sers glamor teetering disdainful slashed trek commandeered inches wide staked by
  reaving sprayed staring uncomfortably embarrassed drip botley's clutching sickeningly
  cotised chunk benerro's took swore backs; nourished fur deceived blessed's planted peat
  slippered talk chores skilled galazza slouching ninth wayward slithering spittle noisy
  deepened dens asha's escaped stood set roar effortless westerlands handed cathterly
  bethany rapes final embroidered goodbrother creighton's talker posted mated ( us grindcorn
   mien poleboat dribbled saddle wrong; quipped angling unknowing boats wouldn't stammered
  stripes cellar restored enormity supported expanse liars mages
```

(4) top-k, k=1:

```
----------- Top-k Sampling 1 -----------
the night is dark and full of terrors . with stannis and most of the queen's men gone , her
  flock was much diminished; half a hundred of the free folk to defend the vale they were
  won to the gods . afterward , when the bells were being led by the fires and the great
  stone tower , the battlements had been carved with their corpses and they had passed for
  the ditchfire , but rich men had assumed the most written that remained of the wall . the
  nights were too small to be away . they had supped on the bare beards of peril , at the
  first sign of a tray . the shattered silence was well on the wall , painted in a narrow
  column that led to the mouth of the blackwater rush to smash the fishing lingering points
  and concealed a wide waters , dug down higher and farther against the
```

(5) top-k, k=20:

```
----------- Top-k Sampling 20 -----------
 the night is dark and full of terrors . with stannis and most of the queen's men gone , her
   flock was much diminished; half a hundred of the free folk to defend the purple cloaks
   that had been the life of old volantis with the tattered prince and the red wedding ,
   their lands and feet of many eyes , for a forge had been gathering in the forest . as for
   the arrival of a host and more of fighting men's lords and captains would pass through the
    streets , one giant and his sons and daughters were children to made the path of courtesy
    as uneasy and red and crumbling . " it is not so long as to breathe his life , " tyrion
   told him . " a victory we all had faced . " " how many five gold you are , " marsh argued
   out over the branches – mussels –
```

(6) top-p, p=0.001:

```
----------- Top-p Sampling 0.001 -----------
 the night is dark and full of terrors . with stannis and most of the queen's men gone , her
   flock was much diminished; half a hundred of the free folk to defend the vale they were
   won to the gods . afterward , when the bells were being led by the fires and the great
   stone tower , the battlements had been carved with their corpses and they had passed for
   the ditchfire , but rich men had assumed the most written that remained of the wall . the
   nights were too small to be away . they had supped on the bare beards of peril , at the
   first sign of a tray . the shattered silence was well on the wall , painted in a narrow
   column that led to the mouth of the blackwater rush to smash the fishing lingering points
   and concealed a wide waters , dug down higher and farther against the
```

(7) top-p, p=0.75:

```
----------- Top-p Sampling 0.75 -----------
 the night is dark and full of terrors . with stannis and most of the queen's men gone , her
   flock was much diminished; half a hundred of the free folk to keep the harvest warm . and
   the night was plain to me ? the war will help the world be fought . they had wells on the
   top of the mountain nonetheless , he almost collided at the air , and for some instant
   they had not tasted all the wildlings . they could be dead , where they had t' learn . all
    the boy … they had seem much close , though . every afternoon he would have given a bite
   of his own , but the great hall was little to pity at sky , as was today . he knew the man
    would give his life . jon had to think . and what would he do ? he found himself resist
```

(8) top-p, p=1:

```
----------- Top-p Sampling 1 -----------
 the night is dark and full of terrors . with him the more , davos realized . " " … what's
   always that they are concerned , it was there was to be a beggar . " " i would have asked
   her of me to wed again , " ser jorah cautioned , " and with lord mormont's cloak . " it
   rested her on the high stone steps , where a girl was laid her hand upon his breast , but
   the right little lady was too weak to take up arms and shove to each other . " you seem to
    find your brother yet as i has ever been before . " " still , " said the old woman on the
    porch . " i do not doubt it . it is the tyrells who killed my father , for this marriage
   . i want you to make me peace . " "
```

Conforming to the question prompt the samples for (2), (4), and (6) are nearly the same due the reason that they focus on nearly the highest probability from the distribution:

(2) Sampling with temperature-scaled $\tau= 0.0001$ would give us a distribution which is peakier so that we avoid the problem of having a long tail of less probable words. Having this τ set to a value as low as 0.0001 helps us in eliminating most of the less probable words while we focus on nearly the highest probability.

(4) Sampling with top-k, k=1 when sorted on probabilities would zero out the probabilities for anything below the 1st token which implies that this would essentially give us the highest probability from the distribution.

(6) In top-p, p=0.001 sampling, we compute the cumulative distribution and cut off as soon as the CDF exceeds p=0.001. Since probabilities are sorted in descending order and p is very low, on a cumulative scale we mean, focus only on highest probabilities.

The samples for (1), (8) are the same due the reason that they sample from the entire distribution:

(1) Vanilla sampling just samples from the entire distribution predicted by the model.

(8) Sampling with top-p, p=1; in terms of CDF, this constitutes the entire distribution and so we are really taking a sample from the entire distribution.

3. Search-based Decoding with Beam Search

1. The `beamsearch()` function that performs beam search until max length was implemented:

```python
def beamsearch(model, text_field, beams=5, prompt="", max_len=50):
    hidden_size = model._modules['rnn'].hidden_size
    n_layers = model._modules['rnn'].num_layers
    h_t_prev = torch.zeros(n_layers, 1, hidden_size)
    c_t_prev = torch.zeros(n_layers, 1, hidden_size)
    w_t = text_field.process([text_field.tokenize(prompt.lower())])
    sampling_criterion = LogSoftmax(dim=1)

    s_t, h_t, c_t = model(w_t, h_t_prev, c_t_prev)
    h_t_b_prev = torch.cat([h_t] * beams, 1)
    c_t_b_prev = torch.cat([c_t] * beams, 1)

    w_t_next = sampling_criterion(s_t[-1])

    topk_vals, topk_indices = torch.topk(w_t_next, beams)
    decoded_strings = topk_indices.view(beams, 1)

    last_prob = []
    for i in range(1, max_len, 1):
        s_t, h_t_b, c_t_b = model(topk_indices, h_t_b_prev, c_t_b_prev)
        w_t_next = sampling_criterion(s_t[-1])

        cumulative_log_probs = w_t_next + topk_vals.view((beams, 1))
        topk_vals, topk_indices = torch.topk(cumulative_log_probs.view(-1), beams)
        b_indices = np.array(np.unravel_index(topk_indices.numpy(), cumulative_log_probs.shape)).T

        h_t_b_new, c_t_b_new = [], []
        for layer in range(n_layers):
            ht_layer, ct_layer = [], []
            for i, j in b_indices:
                ht_layer.append(h_t_b[layer][i])
                ct_layer.append(c_t_b[layer][i])
            h_t_b_new.append(torch.stack(ht_layer))
            c_t_b_new.append(torch.stack(ct_layer))
        h_t_b_prev = torch.stack(h_t_b_new)
        c_t_b_prev = torch.stack(c_t_b_new)

        strings = []
        for i, (r, c) in enumerate(b_indices):
            topk_indices[i] = c
            strings.append(torch.cat([decoded_strings[r], torch.tensor([c])]))

        decoded_strings = strings
        topk_indices = topk_indices.unsqueeze(0)
        last_prob = topk_vals

    decoded_strings = decoded_strings[last_prob.argmax()]
    decoded_strings = prompt + " " + reverseNumeralize(decoded_strings, text_field)
    return decoded_strings
```

2. Outputs for beam search with `b = 1, 10, 50` are as follows:

(1) b = 1:

```
----------- Beam Search B=1 -----------
the night is dark and full of terrors . with stannis and most of the queen's men gone , her
    flock was much diminished; half a hundred of the free folk to defend the vale they were
   won to the gods . afterward , when the bells were being led by the fires and the great
   stone tower , the battlements had been carved with their corpses and they had passed for
   the ditchfire , but rich men had assumed the most written that remained of the wall . the
    nights were too small to be away . they had supped on the bare beards of peril , at the
   first sign of a tray . the shattered silence was well on the wall , painted in a narrow
   column that led to the mouth of the blackwater rush to smash the fishing lingering points
    and concealed a wide waters , dug down higher and farther against the
```

(2) b = 10:

```
----------- Beam Search B=10 -----------
the night is dark and full of terrors . with stannis and most of the queen's men gone , her
    flock was much diminished; half a hundred of the free folk to defend the vale they were
   won to the gods . afterward , when the bells were being led by the fires and the great
   stone tower , the battlements had been carved with their corpses and nurse the eunuch .
   ser jaremy rykker and his lords led them in the river with shields of knights and knights
    , squires , and other merchants rowed a dozen burly khalasars into the outer ward of the
    flint towers , and hundreds of black marble trees were covered with iron – hewn heads of
    ghostskin . " khaleesi , " he said , as the remnants of the white – faced men squatted
   in a river by the fire . " you'd do better here than winterfell , "
```

(3) b = 50:

```
----------- Beam Search B=50 -----------
the night is dark and full of terrors . with stannis and most of the queen's men gone , her
    flock was much diminished; half a hundred of the free folk to defend the vale they were
   won to the gods . afterward , when the bells were being led by the fires and the great
   stone tower , the southron king – beyond – the – wall had not seemed likely there . their
    bodies were smashed and <unk> , and free folk were lit in the flowstone rolling – eyes
   and spears and spears and clams and poles . some of the tents were oats , wheat , the
   faces entering the field , the maps of gargoyles , and the wedding feast , the bright of
   the hair that grew like dangling against his back . when the wind blew forty feet
   streaked , at the center of the wall , the passage grew warmer
```

When the beam width is 1, beam search mimics the behavior of focusing on highest probability in the distribution predicted by the model (same as (2), (4), (6) from the sampling part). This is because we are using top-k sampling in this implementation.

As beam width starts to increase, we notice that the generated samples start to differ at later points in the length.