Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Problem 1

a) Discounted counts under GoodTuring discounting for the three bigrams are:

1. *beer drinker*:

$$\begin{aligned}N^*(w,h) &= [N(w,h)+1]\frac{n_{N(w,h)+1}}{n_{N(w,h)}}\\ &= [1+1]\frac{1600}{5000}\\ &= 0.64\end{aligned}$$

2. *beer lover*:

$$\begin{aligned}N^*(w,h) &= [N(w,h)+1]\frac{n_{N(w,h)+1}}{n_{N(w,h)}}\\ &= [2+1]\frac{800}{1600}\\ &= 1.5\end{aligned}$$

3. *beer glass*:

$$\begin{aligned}N^*(w,h) &= [N(w,h)+1]\frac{n_{N(w,h)+1}}{n_{N(w,h)}}\\ &= [3+1]\frac{500}{800}\\ &= 2.5\end{aligned}$$

b) By using a backing off model, the probabilities of the bigrams are:

1. $p(drinker|beer)$:

$$\begin{aligned}P(w|h) &= \frac{N^*(w,h)}{N(h)} + \lambda(h)\beta(w|h)\\ \lambda(h) &= 1 - \frac{N^*(h)}{N(h)}\\ N(h) &= 6\\ therefore, N^*(h) &= N(h)\\ \lambda(h) &= 0\\ p(drinker|beer) &= \frac{0.64}{6}\\ &= 0.1067\end{aligned}$$

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

2. $p(glass|beer)$:

$$\begin{aligned}
P(w|h) &= \frac{N^*(w,h)}{N(h)} + \lambda(h)\beta(w|h) \\
\lambda(h) &= 0 \\
p(glass|beer) &= \frac{2.5}{6} \\
&= 0.4167
\end{aligned}$$

3. $p(mug|beer)$:

$$\begin{aligned}
P(w|h) &= \frac{N^*(w,h)}{N(h)} + \lambda(h)\beta(w|h) \\
\lambda(h) &= 0 \\
N^*(mug|mug\ \ beer) &= [N(w,h)+1]\frac{n_{N(w,h)+1}}{n_{N(w,h)}} \\
&= [0+1]\frac{N_1}{N_0 N} \\
N_1 &= 5000 \\
N_0 &= 1 \\
N &= \sum_{i=1}^{5} N_i(w,h) n_{N_i(w,h)} \\
N &= 14100 \\
N^*(mug|mug\ \ beer) &= [0+1]\frac{5000}{14100} \\
&= 0.355 \\
p(mug|beer) = [0+1]\frac{0.355}{6} \\
&= 0.059
\end{aligned}$$

Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

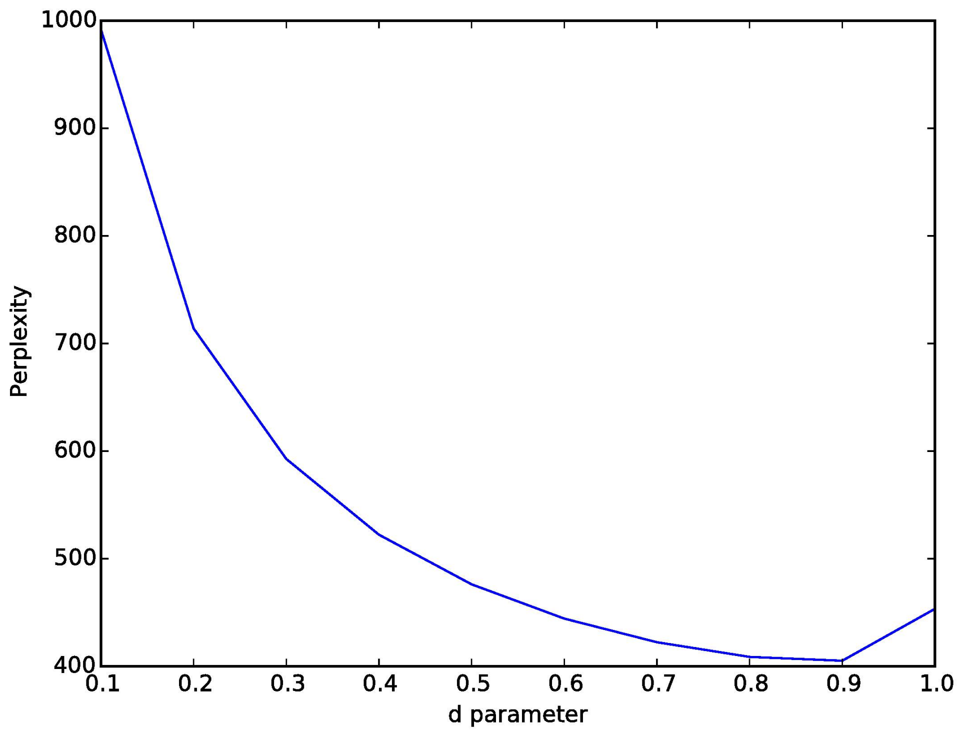
Problem 2

Instruction to run the the code (for Linux):

```
$ python ex5.py
```

We got the following result:

Figure 1:



Ho Vinh Thinh (2562630)
Prabal Agarwal (2561419)
Nurzat Rakhmanberdieva (2561148)

Bonus

1) We have:

$$\begin{aligned}
\frac{N(w,h)+\epsilon}{N(h)+\epsilon V} &= \frac{N(w,h)}{N(h)+\epsilon V} + \frac{\epsilon}{N(h)+\epsilon V} \\
&= \frac{N(w,h)}{N(h)+\epsilon V} * \frac{N(h)}{N(h)} + \frac{\epsilon V}{(N(h)+\epsilon V)V} \\
&= \frac{N(h)}{N(h)+\epsilon V} * \frac{N(w,h)}{N(h)} + \frac{N(h)-N(h)+\epsilon V}{N(h)+\epsilon V} * \frac{1}{V} \\
&= \frac{N(h)}{N(h)+\epsilon V} * \frac{N(w,h)}{N(h)} + \left(\frac{N(h)+\epsilon V}{N(h)+\epsilon V} - \frac{N(h)}{N(h)+\epsilon V}\right) * \frac{1}{V} \\
&= \frac{N(h)}{N(h)+\epsilon V} * \frac{N(w,h)}{N(h)} + \left(1 - \frac{N(h)}{N(h)+\epsilon V}\right) * \frac{1}{V}
\end{aligned}$$

With $\mu = \dfrac{N(h)}{N(h)+\epsilon V}$, then we have:

$$\frac{N(w,h)+\epsilon}{N(h)+\epsilon V} = \mu * \frac{N(w,h)}{N(h)} + (1-\mu) * \frac{1}{V}$$

2) In case if the probability of a word estimated by a unigram model is 0, means that the word is an OOV word. Instead of using a zerogram model probability value, we could improve it by adding the average probability value of the words similar to the current word. Hence metric will be:

$$\begin{aligned}
P(w) &= \frac{N(w)}{N}, if N(w) > 0 \\
&= (1-\lambda)\left(\sum_{v \in Syn(w)} \frac{N(v)}{N}\right) + \lambda c; N(w) = 0
\end{aligned}$$

Syn(w) is a set of terms similar to w and c is the probability value obtained from zero gram model.