Natural Language Processing - Problem Set #4

Emma Ziegellaub Eichler - edz2103@columbia.edu

November 30, 2012

Problem 1

a

Since $f_1(x,y) = f_2(x,y)$, we have either $f(x,y) = (0,0)$ or $f(x,y) = (1,1)$. Let k be the number of examples in the training sample such that $f(x,y) = (1,1)$. Then $n-k$ is the number of examples in the training sample such that $f(x,y) = (0,0)$. Consider the case where $v' = (v,v)$. Then $L(v') = \Sigma_i log P(y_i|x_i, v) - c\Sigma_k v_k^2 = L(v) = \Sigma_i log \frac{e^{v' \cdot f(x,y)}}{\Sigma_{y' \in V} e^{v \cdot f(x,y')}} = klog\frac{2^{2v}}{ke^{2v}+n-k} + (n-k)log\frac{1}{ke^{2v}+n-k} - 2cv^2 = 2kv - klog(ke^{2v} + n - k) + (n-k)log(1) + (k-n)log(ke^{2v}+n-k) - 2cv^2 = 2kv - nlog(ke^{2v}+n-k) - 2cv^2$. Now consider the case where $v'' = (v_1, v_2)$, $v_1 \neq v_2$. Following the same simplification steps as before, we have $L(v'') = k(v_1+v_2) - nlog(ke^{v_1+v_2} + n - k) - c(v_1^2 + v_2^2)$. Now to show v^* satisfies $v_1 = v_2$, we need only fix $v_1 + v_2 = 2v$ and show $L(v') > L(v'')$. Since $v_1 + v_2 = 2v$, $2kv - nlog(ke^{2v} + n - k) = k(v_1 + v_2) - nlog(ke^{v_1+v_2} + n - k)$, so we only need to show $2v^2 < v_1^2 + v_2^2$. Let $v_1 = v + \alpha$ so $v_2 = v - \alpha$. Then $v_1^2 + v_2^2 = (v+\alpha)^2 + (v - \alpha^2) = 2v^2 - 2v\alpha + 2v\alpha + 2\alpha^2 = 2v^2 + 2\alpha^2$. Since $|\alpha| > 0$, $\alpha^2 > 0$, so $2v^2 < 2v^2 + 2\alpha^2$ and therefore $L(v') > L(v'')$ so v^* satisfies $v_1^* = v_2^*$. $\square$

b

Now we define $L(v) = \Sigma_i log P(y_i|x_i, v) - c\Sigma_k |v_k|$. As before, we let k be the number of examples in the training sample such that $f(x,y) = (1,1)$, so $n-k$ is the number of examples in the training sample such that $f(x,y) = (0,0)$. Again, we evaluate $L(v')$ and $L(v'')$ where $v' = (v,v)$ and $v'' = (v_1, v_2)$, $v_1 \neq v_2$. Following the same simplification steps as in the previous part, we have $L(v') = 2kv - nlog(ke^{2v} + n - k) - 2c|v|$ and $L(v'') = k(v_1 + v_2) - nlog(ke^{v_1+v_2} + n - k) - c(|v_1| + |v_2|)$. Again, fixing $v_1 + v_2 = 2v$, $2kv - nlog(ke^{2v}+n-k) = k(v_1+v_2) - nlog(ke^{v_1+v_2}+n-k)$, so to find $max\{L(v'), L(v'')\}$, we need only find $min\{2|v|, |v_1| + |v_2|\}$. As before, let $v_1 = v + \alpha$ so $v_2 = v - \alpha$. For $-\alpha < v < \alpha$, $|2v| < |v+\alpha| + |v - \alpha| = |v_1| + |v_2|$, and for $-v \leq \alpha \leq v$, $|2v| = |v+\alpha| + |v-\alpha| = |v_1| + |v_2|$. Thus, v^* satisfies $-v_1^* - v_2^* \leq v_1^* - v_2^* \leq v_1^* + v_2^*$. $\square$

2

Since $v^* = argmax_v L(v)$, $L'(v^*) = 0$. We know $\frac{dL(v)}{dv_k} = \Sigma_{i=1}^n f_k(x_i, y_i) - \Sigma_{i=1}^n \Sigma_{y \in Y} P(y|x_i, v^*) f_k(x_i, y)$, so we set $\Sigma_{i=1}^n f_k(x_i, y_i) - \Sigma_{i=1}^n \Sigma_{y \in Y} P(y|x_i, v^*) f_k(x_i, y) = 0$. Since $f_k(x_i, y_i) = 1$ if and only if $x_i = x_k$ and $y_i = y_k$ (and $f_k(x_i, y_i) = 0$ otherwise), $\Sigma_{i=1}^n f_k(x_i, y_i) = Count(x_k, y_k)$. Since $\Sigma_{y \in Y} P(y|x_i, v^*) f_k(x_i, y) = 0$ if $x_i \neq x_k$ and $\Sigma_{y \in Y} P(y|x_i, v^*) f_k(x_i, y) = \Sigma_{y \in Y} P(y|x_k, v^*) f_k(x_k, y)$ if $x_i = x_k$, $\Sigma_{i=1}^n \Sigma_{y \in Y} P(y|x_i, v^*) f_k(x_i, y) = Count(x_k) \Sigma_{y \in Y} P(y|x_k, v^*) f_k(x_k, y)$. If we take $P(x|y, v) = \frac{Count(x.y)}{Count(x)}$, we have $Count(x_k) \Sigma_{y \in Y} P(y|x_k, v^*) f_k(x_k, y) = Count(x_k) \Sigma_{y \in Y} \frac{Count(x_k,y)}{Count(x_k)}$ $Count(x_k) \frac{1}{Count(x_k)} \Sigma_{y \in Y} Count(x_k, y) f_k(x_k, y) = \Sigma_{y \in Y} Count(x_k, y) f_k(x_k, y)$. Since $f_k(x_k, y) = 1$ if $y = y_k$ and $f_k(x_k, y) = 0$ if $y \neq y_k$, $\Sigma_{y \in Y} Count(x_k, y) f_k(x_k, y) = Count(x_k, y_k)$. Then $\frac{dL(v)}{dv_k} = Count(x_k, y_k) - Count(x_k, y_k) = 0$. Since $L(v)$ is concave, it has one maximum and no minimum, so we have found the global maximum, v^*, and therefore it must satisfy $P(x|y, v^*) = \frac{Count(x,y)}{Count(x)}$. $\square$

3

a

We choose the log-linear model with inputs $X = V$, labels $Y = V \bigcup V'$, $d = 2$, function $f : XxY \to \mathsf{R}^2$ where $f(x,y) = (f_1(x,y), f_2(x,y))$, $f_1(x,y) = 1$ if $x = y$ and $f_1(x,y) = 0$ if $x \neq y$, and $f_2(x,y) = 1$ if $x = y'$ and $f_2(x,y) = 0$ if $x \neq y'$.

b

Using model defined above, $f(x,x) = (1,0)$, $f(x,x') = (0,1)$, $f(x,y) = (0,0)$ where $y \notin \{x, x'\}$. Then we also have $\Sigma_{y \in Y} e^{(v_1,v_2)\cdot f(x,y)} = (v_1,v_2)\cdot(0,0) + e^{(v_1,v_2)\cdot(1,0)} + e^{(v_1,v_2)\cdot(0,1)}$, $\forall x \in X$ where the parameter vector $v = (v_1, v_2)$. Using these substitutions and the probabilities given in the problem, we have the following equations:
$P(the|the) = \frac{e^{(v_1,v_2)\cdot(1,0)}}{(|V'|-2)e^{(v_1,v_2)\cdot(0,0)}+e^{(v_1,v_2)\cdot(1,0)}+e^{(v_1,v_2)\cdot(0,1)}} = 0.4$
$P(eht|the) = \frac{e^{(v_1,v_2)\cdot(0,1)}}{(|V'|-2)e^{(v_1,v_2)\cdot(0,0)}+e^{(v_1,v_2)\cdot(1,0)}+e^{(v_1,v_2)\cdot(0,1)}} = 0.3$
$P(dog|the) = \frac{e^{(v_1,v_2)\cdot(0,0)}}{(|V'|-2)e^{(v_1,v_2)\cdot(0,0)}+e^{(v_1,v_2)\cdot(1,0)}+e^{(v_1,v_2)\cdot(0,1)}} = \frac{0.3}{|V'|-2}$

c

Using the equations from the previous part, we evaluate $e^{(v_1,v_2)\cdot(0,0)} = e^0 = 1$, $e^{(v_1,v_2)\cdot(1,0)} = e^{v_1}$, $e^{(v_1,v_2)\cdot(0,1)} = e^{v_2}$ and substitute, giving us the following equations:
$\frac{1}{|V'|-2+e^{v_1}+e^{v_2}} = \frac{3}{10(|V'|-2)}$
$\frac{e^{v_2}}{|V'|-2+e^{v_1}+e^{v_2}} = \frac{3}{10}$
$\frac{e^{v_2}}{|V'|-2+e^{v_1}+e^{v_2}} = \frac{4}{10}$
Then $|V'| - 2 + e^{v_1} + e^{v_2} = \frac{10(|V'|-2)}{3}$, so solving for v_1 and v_2 and substituting this in yields: $v_2 = ln(|V'|-2)$ and $v_1 = ln(\frac{4}{3}) + ln(|V'|-2)$.
Then our parameter vector $v = (v_1, v_2) = (ln(\frac{4}{3}) + ln(|V'|-2), ln(|V'|-2))$.