LECTURE 9

Last time:

- Channel capacity
- Binary symmetric channels
- Erasure channels
- Maximizing capacity

Lecture outline

- Maximizing capacity: Arimoto-Blahut
- Convergence
- Examples

Arimoto-Blahut

Lemma 1:

$$I(X;Y) = \max_{\widehat{P}_{X|Y}} \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x) \log\left(\frac{\widehat{P}_{X|Y}(x|y)}{P_X(x)}\right)$$

Proof:

$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{x\in\mathcal{Y}} P_{X|Y}(x|y) P_Y(y) \log\left(\frac{P_{X|Y}(x|y)}{P_X(x)}\right)$

Recall:

$P_{X|Y}(x|y) = \frac{P_X(x)P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X(x')P_{Y|X}(y|x')}$

and

$P_Y(y) = \sum_{x'\in\mathcal{X}} P_X(x') P_{Y|X}(y|x')$

Arimoto-Blahut

$$
\begin{aligned}
& I(X;Y) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x) \\
& \log\left(\frac{\widehat{P}_{X|Y}(x|y)}{P_X(x)}\right) \\
= \; & I(X;Y) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_Y(y) P_{X|Y}(x|y) \\
& \log\left(\frac{\widehat{P}_{X|Y}(x|y)}{P_X(x)}\right) \\
= \; & \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_Y(y) P_{X|Y}(x|y) \log\left(\frac{P_{X|Y}(x|y)}{\widehat{P}_{X|Y}(x|y)}\right) \\
& \text{(using } \log(x) \geq 1 - \tfrac{1}{x}\text{)} \\
\geq \; & \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_Y(y) P_{X|Y}(x|y) \\
- \; & \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_Y(y) \widehat{P}_{X|Y}(x|y) \\
= \; & 0
\end{aligned}
$$

Arimoto-Blahut

Capacity is

$$C = \max_{P_X} \max_{\widehat{P}_{X|Y}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x) \log\left(\frac{\widehat{P}_{X|Y}(x|y)}{P_X(x)}\right)$$

For fixed P_X, RHS is maximized when

$$\widehat{P}_{X|Y}(x|y) = \frac{P_X(x) P_{Y|X}(y|x)}{\sum_{x' \in \mathcal{X}} P_X(x') P_{Y|X}(y|x')}$$

For fixed $\widehat{P}_{X|Y}$, RHS is maximized when

$$P_X(x) = \frac{e^{\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log(\widehat{P}_{X|Y}(x|y))}}{\sum_{x' \in \mathcal{X}} \left(e^{\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x') \log(\widehat{P}_{X|Y}(x'|y))}\right)}$$

Arimoto-Blahut

Combining the two means maximization when

$$P_X(x) = \frac{e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)\,\log(\widehat{P}_{X|Y}(x|y))}}{\sum_{x'\in\mathcal{X}}\left(e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x')\,\log(\widehat{P}_{X|Y}(x'|y))}\right)}$$

$$= \frac{P_X(x)e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)\,\log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X(x')P_{Y|X}(y|x')}\right)}}{\sum_{x'\in\mathcal{X}} P_X(x')\left(e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x')\,\log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X(x')P_{Y|X}(y|x')}\right)}\right)}$$

Note also that $\sum_{x\in\mathcal{X}} P_X(x) = 1$.

This may be very hard to solve.

Arimoto-Blahut

Proof:

The first two statements follow immediately from our lemma

For any value of x where $P_{X|Y}(x|y) = 0$, $P_X(x)$ should be set to 0 to obtain the maximum.

To find the maximum over the PMF P_X, let us first ignore the constraint of positivity and use a Lagrange multiplier for the $\sum_x P_X(x) = 1$

Then

$$\frac{\partial}{\partial P_X(x)}\{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)P_{Y|X}(y|x)\log\left(\frac{\widehat{P}_{X|Y}(x|y)}{P_X(x)}\right)+ \lambda\left(\sum_{x\in\mathcal{X}} P_X(x)-1\right)\} = 0$$

Arimoto-Blahut

Equivalently

$$-\log(P_X(x)) - 1 + \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log\left(\widehat{P}_{X|Y}(x|y)\right) + \lambda = 0$$

so

$$P_X(x) = \frac{e^{\sum_{x \in \mathcal{X}} P_{Y|X}(y|x) \log\left(\widehat{P}_{X|Y}(x|y)\right)}}{\sum_{x \in \mathcal{X}} e^{\sum_{x \in \mathcal{X}} P_{Y|X}(y|x) \log\left(\widehat{P}_{X|Y}(x|y)\right)}}$$

(this ensures that λ is such that the sum of the $P_X(x)$s is 1)

What about the constraint we did not use for positivity?

The solution we found satisfies that.

Convergence of Arimoto-Blahut

Let P_X^0 be a PMF and let

$$P_X^{r+1}(x) = P_X^r(x) \frac{c_x\left(P_X^r(x_1),\ldots,P_X^r(x_{|\mathcal{X}|})\right)}{\sum_{x' \in \mathcal{X}} c_x\left(P_X^r(x_1),\ldots,P_X^r(x_{|\mathcal{X}|})\right) P_X^r(x')}$$

where

$$c_x\left(P_X^r(x_1),\ldots,P_X^r(x_{|\mathcal{X}|})\right)$$

$$= e^{\sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log\left(\frac{P_{Y|X}(y|x)}{\sum_{x' \in \mathcal{X}} P_X(x') P_{Y|X}(y|x')}\right)}$$

the sequence I^r of $I(X;Y)$ for X taking the PMF P_X^R for I^r converges to C from below

Convergence of Arimoto-Blahut

Proof:

For any given P_X^r, we can increase mutual information by taking

$$P_{Y|X}^r = \frac{P_X^r(x)P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X^r(x')P_{Y|X}(y|x')}$$

With $P_{Y|X}^r$ fixed, then choose P_X^{r+1} by

$$P_X^{r+1}(x) = \frac{e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)\log(P_{X|Y}^r(x|y))}}{\sum_{x'\in\mathcal{X}} e^{\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x')\log(P_{X|Y}^r(x'|y))}}$$

If we define

$$J^r = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X^{r+1}(x)P_{Y|X}(y|x)\log\left(\frac{P_{X|Y}^r(x|y)}{P_X^{r+1}(x)}\right)$$

Then $I^r \leq J^r \leq I^{r+1} \leq J^{r+1} \leq \ldots$

This an upper bounded non-decreasing sequence, therefore it reaches a limit

Convergence of Arimoto-Blahut

Why is the limit C?

Let P_X^* be a capacity achieving PMF

$$\sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^{r+1}(x)}{P_X^r(x)}\right)$$

$$= \sum_{x\in\mathcal{X}} P_X^*(x)$$

$$\log\left(\frac{c_x\left(P_X^r(x_1),...,P_X^r(x_{|\mathcal{X}|})\right)}{\sum_{x'\in\mathcal{X}} c_x\left(P_X^r(x_1),...,P_X^r(x_{|\mathcal{X}|})\right) P_X^r(x')}\right)$$

$$= \sum_{x\in\mathcal{X}} P_X^*(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)$$

$$\log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X^r(x') P_{Y|X}(y|x')}\right)$$

$$- \sum_{x\in\mathcal{X}} P_X^*(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x)$$

$$\log\left(\sum_{x'\in\mathcal{X}} P_X^r(x')\right.$$

$$\left. e^{\sum_{y'\in\mathcal{Y}} P_{Y|X}(y'|x') \log\left(\frac{P_{Y|X}(y'|x')}{\sum_{x''\in\mathcal{X}} P_X^r(x'') P_{Y|X}(y'|x'')}\right)}\right)$$

Convergence of Arimoto-Blahut

By considering the K-L distance, we have that

$\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X^*(x)P_{Y|X}(y|x)$

$\log\left(\frac{\sum_{x'\in\mathcal{X}} P_X^*(x')P_{Y|X}(y|x')}{\sum_{x'\in\mathcal{X}} P_X^r(x')P_{Y|X}(y|x')}\right) \geq 0$

so

$$
\begin{aligned}
& \sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^{r+1}(x)}{P_X^r(x)}\right) \\
\geq\ & \sum_{x\in\mathcal{X}} P_X^*(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) \\
& \log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X^*(x')P_{Y|X}(y|x')}\right) \\
-\ & \sum_{x\in\mathcal{X}} P_X^*(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) \\
& \log\left(\sum_{x'\in\mathcal{X}} P_X^r(x') \right. \\
& \left. e^{\sum_{y'\in\mathcal{Y}} P_{Y|X}(y'|x') \log\left(\frac{P_{Y|X}(y'|x')}{\sum_{x''\in\mathcal{X}} P_X^r(x'')P_{Y|X}(y'|x'')}\right)}\right)
\end{aligned}
$$

Convergence of Arimoto-Blahut

Hence

$$\begin{aligned}
&\sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^{r+1}(x)}{P_X^r(x)}\right)\\
\geq\ & C - J^r
\end{aligned}$$

Sum over r

$$\begin{aligned}
&\sum_{r=0}^{m} (C - J^r)\\
\leq\ & \sum_{r=0}^{m} \sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^{r+1}(x)}{P_X^r(x)}\right)\\
=\ & \sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^{m+1}(x)}{P_X^0(x)}\right)\\
\leq\ & \sum_{x\in\mathcal{X}} P_X^*(x) \log\left(\frac{P_X^*(x)}{P_X^0(x)}\right)
\end{aligned}$$

$C - J^r \geq 0$ and non increasing, with bounded sum, so it goes to 0, hence J^r converges to C

In practice, **convergence can be very slow**

Example

Other types of maximization

Interior point methods

Cutting plane algorithms