Sparse Regression Codes

Сенин Игорь

Московский Физико-Технический Институт

senin.ia@phystech.edu

17 мая 2024 г.

Предыстория и литература

- A. Barron, A. Joseph: "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity"(2012)
- A. Joseph, A. R. Barron: "Fast sparse superposition codes have near exponential error probability for $R < C$"(2014)
- R. Venkataramanan, T. Sarkar, S. Tatikonda: "Lossy compression via sparse linear regression: Computationally efficient encoding and decoding"(2014)
- T. M. Cover, J. A. Thomas: "Elements of Information Theory"

Цели и задачи

- Изучить SPARC
- Повторить результаты статей
- Написать что-то новое

Теория

Теорема Шеннона-Хартли для AWGN канала

$\mathcal{C}$ - точная верхняя грань всех достижимых R.

Теорема Шеннона-Хартли для lossy compression

$R(D) = R^I(D) = \inf\limits_{p(\hat{x}|x)} I(X;\hat{X})$, где инфимум берётся по всем распределениям $p(\hat{x}|x)$, что $\sum\limits_{(x,\hat{x})} p(x,\hat{x}) \cdot d(x,\hat{x}) \leqslant D$.

Теория

Канал с шумом

Цель - придумать код с быстрыми алгоритмами кодирования и декодирования, который (асимптотически) достигал бы пропускной способности гауссовского канала $\frac{1}{2}\log(1 + snr)$, $snr = \frac{P}{\sigma^2}$.

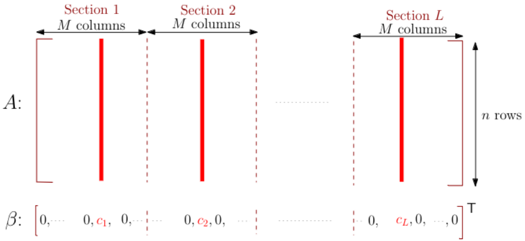
Сжатие с потерями

Аналогично, цель - придумать код с быстрыми алгоритмами кодирования и декодирования, который достигал бы rate-distortion функции гауссовского источника $\frac{1}{2}\log\frac{\sigma^2}{D}$.

Актуальность

- Передача информации в беспроводных средах: спутниковые и мобильные сети
- Восстановление двоичных сигналов
- Сжатие с непрерывным алфавитом

В SPARC'ах кодовые слова - вектора длины n вида $A\beta$, где A - матрица плана n x ML, β - разреженный вектор длины ML.
A и β разбиты на L секций длины M, причём в каждой из L секций β ровно одна ненулевая координата c_l.



Параметры M, L, элементы A, c_1, ..., c_L фиксируются и считаются известными.

SPARCs

- Элементы матрицы A выбираются i.i.d. $\mathcal{N}(0, \frac{1}{n})$.
- L выбирается $\Theta(n/\log n)$.
- $M = L^a$ для некоторой константы $a > 0$.

Благодаря такому выбору M и L размер матрицы A растёт полиномиально с n.

Общее число кодовых слов - M^L.
$M^L = 2^{Rn} \Leftrightarrow L \log M = Rn$

SPARCs

Оптимальный энкодер

Пусть $s = (s_1, ..., s_n)$ - набор вещественных чисел, которые хотим закодировать.

- Энкодер: $\hat{\beta} = \text{argmin}||s - A\beta||^2$.
- Декодер: $\hat{s} = A\hat{\beta}$.

Кодовое слово - набор индексов ненулевых элементов в векторе $\hat{\beta}$, то есть $L \log M$ бит. На практике количество бит уменьшается в $\log n$ раз.

Successive-approximation энкодер

Идея: итеративно аппроксимировать остатки вида $\mathcal{R}_i = \mathcal{R}_{i-1} - c_i A_{m_i}$; $R_0 = s$.

Числа c_i - специального вида, $c_i = \sqrt{\frac{2R\sigma^2}{L}(1 - \frac{2R}{L})^{i-1}}$.

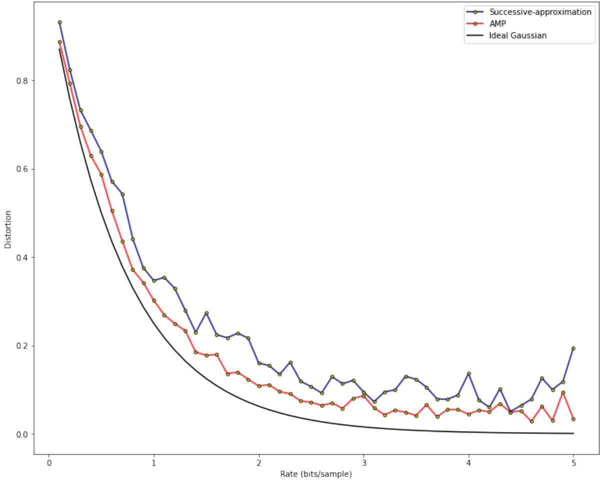
AMP энкодер

За заданное число итераций находить лучшее приближение задачи Lasso: $\hat{\beta} = argmin||y - A\beta||_2^2 + \lambda||\beta||_1$ для заданного λ.

Результаты

- Изучена теория вокруг SPARC
- Написан мини-фреймворк для сжатия с потерями на основе SPARC
- Проведено сравнение двух энкодеров

Сравнение Successive и AMP энкодеров



Что осталось сделать

- Отчитаться перед руководителем
- Привести фреймворк к хорошему виду
- Выложить фреймворк и сравнение в открытый доступ