Machine Learning Talk VI

Matrix Completion and Sparse Recovery

Axel G. R. Turnquist

NJIT Department of Mathematical Sciences

October 28, 2020

Best rank-k approximation

Take an n by m matrix A, with rank r. Suppose we wish to find the matrix of rank $k \leq r$, which we call A_k such that:

$$A_k = \text{argmin}_B ||A - B||_2 \tag{1}$$

Eckart-Young-Mirsky Theorem:

$$A_k = \sum_{i=1}^{k} s_i(A) u_i v_i^T \tag{2}$$

which is the singular value decomposition, with all singular values above s_k set to zero.

Frobenius Norm

$$||A||_F = \left(\sum_i \sum_j |A_{ij}|^2 \right)^{1/2} \tag{3}$$

Defines an inner product!

$$\langle A, B \rangle_F = \sum_i \sum_j A_{ij} B_{ij} \tag{4}$$

Clearer representation:

$$||A||_F = \left(\sum_{i=1}^{r} s_i(A)^2 \right)^{1/2} \tag{5}$$

Frobenius vs. Operator Norm

The following shows the relation between the norms:

$$||A||_2 \leq ||A||_F \leq \sqrt{r}||A||_2 \tag{6}$$

Or:

$$1 \leq \frac{||A||_F^2}{||A||_2^2} \leq r \tag{7}$$

The middle term is known as the **stable rank** or statistical rank. Robust to perturbations.

Matrix Approximation/Completion

Suppose we are only shown random entries of an $n \times n$ matrix X, with rank r. More specifically, we are shown Y, where:

$$Y_{ij} = \delta_{ij} X_{ij}, \text{ where } \delta_{ij} \sim \text{Ber}(p) \text{ are indep.} \tag{8}$$

On average we are shown m entries of X. Let $\hat{X}$ be a best-r-rank approximation of $p^{-1}Y$. Then,

$$\mathbb{E}\frac{1}{n}||\hat{X} - X||_F \leq C\sqrt{\frac{rn\log n}{m}}||X||_\infty \tag{9}$$

Observation: for $r \ll n \log n$, then matrix completion is possible. Matrix completion is not bad for a low-rank matrix!

Sparsity

Two sagacious observations:

- "We are drowning in information and starving for knowledge." — *Rutherford Roger*
- "Use a procedure that does well in sparse problems, since no procedure does well in dense problems." — *Applied math adage*

Notions of sparsity:

1. Small stable rank
2. Few parameters relevant to the model (many zeroes). This notion is better for large calculations.

Solving a Problem

There are different increasingly weaker notions of what we mean to solve a problem with a solution x:

1. **Strong**: Find $\hat{x}$ such that $x = \hat{x}$ (a.e. or everywhere)
2. **Weaker**: Find $\hat{x}$ such that $||\hat{x} - x|| \leq$ small controllable amount
3. **Weakest**: Find $\hat{x}$ such that:

$$\mathbb{E}||\hat{x} - x|| \leq \text{small controllable amount} \tag{10}$$

$\iff$ in concentration settings, the event $||\hat{x} - x|| \leq$ small controllable amount happens with high probability.

Sparse Recovery

Suppose we have a signal x, with linear measurements of x. We can represent this as:

$$y = Ax \tag{11}$$

where y and A, an $m \times n$ matrix are known. Can we recover x? Sometimes this is easy. But, suppose we have $m \ll n$. This is **ill-posed**. Should we give up? In data science some problems have these dimensions. How to get close to the solution x?

1. Know that the entries, or rows of A have nice properties (subgaussian, indep., isotropic)
2. Incorporate prior information $x \in T$, hope the set T is small

Recovery

Start by finding any $\hat{x}$ that solves the equation, i.e. $y = A\hat{x}$ and $\hat{x} \in T$, we can guarantee:

$$\mathbb{E}||\hat{x} - x||_2 \leq \frac{CK^2 w(T)}{\sqrt{m}} \tag{12}$$

Suppose we measure $m \geq C(K^4/\epsilon^2)d(T)$ times. Then,

$$\mathbb{E}||\hat{x} - x||_2 \leq \epsilon \text{diam}(T) \tag{13}$$

Notice that when $n \gg C(K^4/\epsilon^2)d(T)$, we can approximately solve the ill-posed problem we wanted to abandon earlier!

The Sparsity of the $||\cdot||_1$-norm

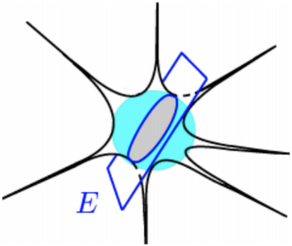
Recall that the set:

$$B_1^n := \{v : ||v||_1 \leq 1\} \tag{14}$$

is convex, yet spiky (in the sense of my previous talk). Vectors sampled randomly from this set are asymptotically close to the origin as $n \to \infty$. That is,

$$\mathbb{E}(||v||_2) \sim \frac{\log n}{n} \tag{15}$$

Hence, imposing this prior, not only is a convex problem, but also leads to sparse data.


E

Solving a Sparse Problem Using the $||\cdot||_1$-norm

Suppose that we know that $||x||_0 = s$, i.e. there are known to be less than or equal to s non-zero entries. Can we solve:

$$y = Ax, \quad ||x||_0 \leq s \tag{16}$$

Hard problem in practice. Can we approximate by a convex problem? Yes,

$$\text{minimize } ||x'||_1 \text{ s.t. } y = Ax' \tag{17}$$

Use known algorithms to find a solution $\hat{x}$. Then,

$$\mathbb{E}||\hat{x} - x||_2 \leq CK^2\sqrt{\frac{s \log n}{m}} \tag{18}$$

Exact Recovery

In some cases it is highly likely the recovery is exact!

$$\text{minimize } ||x'||_1 \text{ s.t. } y = Ax' \quad (19)$$

Solve the optimization problem to get a minimizer $\hat{x}$.

The event below happens with probability at least

$$1 - 2\exp(-cm/K^4) \quad (20)$$

Assume x is s-sparse and the number of measurements satisfies $m \geq CK^4 s \log n$. Then a solution $\hat{x}$ of the convex optimization problem is exact!!!!!!!!

$$\hat{x} = x \quad (21)$$

Questions?

Highlighted Resources

- “**High-Dimensional Probability**” Vershynin, Roman.
- “**Statistical Learning with Sparsity: The Lasso and Generalizations**” Trevor Hastie, Robert Tibshirani, Martin Wainwright.
- “**Foundations of Machine Learning**” Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar

Future Talks

Further potential topics:

- Adversarial attacks

Nov. 6: Yixuan Sun GAN/WGAN