8. Variations on least-squares

- ℓ_2 regularization
- ℓ_1 regularization
- Robust least-squares
- Least-squares with constraints

ℓ_2-regularized least-squares

$$\underset{x}{\text{minimize}} \quad \|Ax - y\|_2^2 + \gamma \|x\|_2^2$$

- provides an approximate solution of $Ax \approx y$ with *minimum-norm* x
- also called *Tikhonov regularized least-squares*
- $\gamma > 0$ controls the trade off between the fitting error and the size of x
- has the analytical solution for any $\gamma > 0$:

$$x = (A^*A + \gamma I)^{-1} A^* y$$

  (no restrictions on shape, rank of A)
- interpreted as a MAP estimation with the log-prior of the Gaussian

ℓ_1-regularized least-squares

Idea: adding $|x|$ to a minimization problem introduces a sparse solution

Consider a scalar problem:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}(x-a)^2 + \gamma|x|$$

The optimal solution is

$$x^* = \begin{cases} (|a|-\gamma)\,\mathbf{sign}(a), & |a| > \gamma \\ 0, & |a| \leq \gamma \end{cases}$$

If γ is large enough, x^* will be zero

For $x \in \mathbf{R}^n$, recall

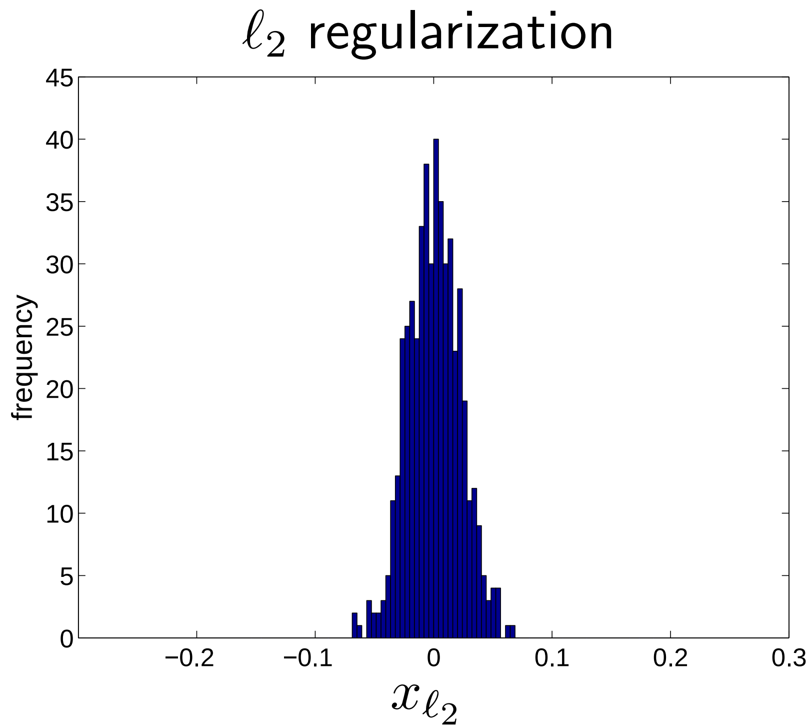
$$\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

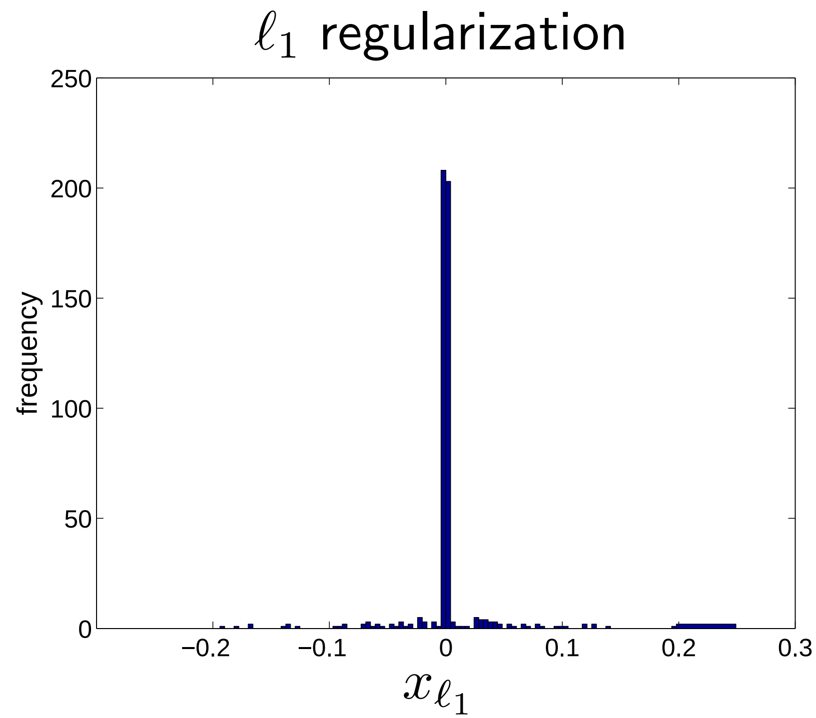
Extend this idea by adding the ℓ_1-norm penalty to the least-square problem

$$\underset{x}{\text{minimize}} \quad \|Ax - y\|_2^2 + \gamma \|x\|_1$$

- a convex heuristic method for finding a sparse x that gives $Ax \approx y$
- also called *Lasso* or *basis pursuit*
- no analytical solution, but can be solved efficiently
- interpreted as a MAP estimation with the log-prior of the Laplacian distribution

Example $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ with $m = 100, n = 500, \gamma = 0.2$





- The solution of ℓ_2 regularization is more widely spread
- The solution of ℓ_1 regularization is concentrated at zero

Robust least-squares

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|$$

but A may have variation or some uncertainty

Worst-case robust least-squares

Describe the uncertainty by a set of possible values for A:

$$A \in \mathcal{A} \subseteq \mathbf{R}^{m \times n}$$

The problem is to minimize the worst-case error:

$$\underset{x}{\text{minimize}} \quad \sup_{A} \ \{\|Ax - y\| \mid A \in \mathcal{A}\}$$

- always a convex problem
- its tractablity depends on the norm used and the description of $\mathcal{A}$

Stochastic robust least-squares

When A is a random variable, so we can describe A as

$$A = \bar{A} + U,$$

where $\bar{A}$ is the average value of A and U is a random matrix

Use the expected value of $\|Ax - y\|$ as the objective:

$$\underset{x}{\text{minimize}} \quad \mathbf{E}\,\|Ax - y\|_2^2$$

Expanding the objective gives

$$\begin{aligned}\mathbf{E}\,\|Ax - y\|_2^2 &= (\bar{A}x - y)^*(\bar{A}x - y) + \mathbf{E}\,x^*U^*Ux \\ &= \|\bar{A}x - y\|_2^2 + x^*Px\end{aligned}$$

where $P = \mathbf{E}\,U^*U$

This problem is equivalent to

$$\underset{x}{\text{minimize}} \quad \|\bar{A}x - y\|_2^2 + \|P^{1/2}x\|_2^2$$

with solution

$$x = (\bar{A}^* A + P)^{-1} \bar{A}^* y$$

- a form of a regularized least-squares
- balance making $\bar{A}x - b$ small with the desire for a small x (so that the variation in Ax is small)
- Tikhonov regularization is a special case of robust least-squares:

  when U has zero mean and uncorrelated variables, $i.e.$, $\mathbf{E}\, U^*U = \delta I$

Least-squares with constraints

$$\begin{array}{ll} \text{minimize} & \|Ax - y\| \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

$\mathcal{C}$ is a convex set

- used to rule out certain unacceptable approximations of y
- arise as prior knowledge of the vector x to be estimated
- same as determining the projection of y on a set more complicated than a subspace
- form a convex optimization problem with no analytical solution (typically)

Nonnegativity constraints on variables

$$\mathcal{C} = \{ x \mid x \succeq 0 \}$$

- parameter x known to be nonnegative, e.g., powers, rates, etc.
- finding the projection of y onto the *cone* generated by the columns of A

Variable bounds

$$\mathcal{C} = \{ x \mid l \preceq x \preceq u \}$$

- vector x known to lie in an interval $[l, u]$
- finding the projection of y onto the image of a box under the linear mapping induced by A

Probability distribution

$$\mathcal{C} = \{ \, x \mid x \succeq 0, \quad \mathbf{1}^T x = 1 \, \}$$

- arise in estimation of proportions which are nonnegative and sum to one
- approximating y by a convex combination of the columns of A

Norm ball constraint

$$\mathcal{C} = \{ \, x \mid \|x - x_0\| \leq d \, \}$$

where x_0 and d are problem parameters

- x_0 is a prior guess of what x should be
- d is the maximum plausible deviation from our prior guess
- the constraints $\|x - x_0\| \leq d$ can denote a *trust region*. (The linear relation $y = Ax$ is an approximation and only valid when x is near x_0)

References

Chapter 4 in
T. Söderström and P. Stoica, *System Identification*, Prentice Hall, 1989

Chapter 2-3 in
T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, 2000

Chapter 6 in
S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge press, 2004