Introduction

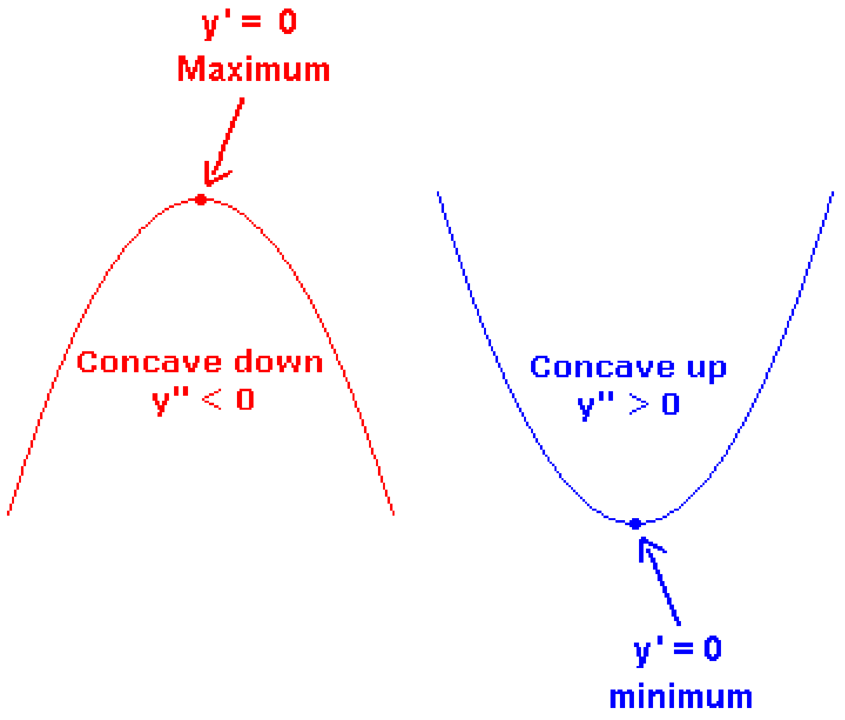
- Convex optimization
- Sparsity
- Machine Learning: Dictionary Learning

Optimization

- Optimization
  - Max/Min over a set

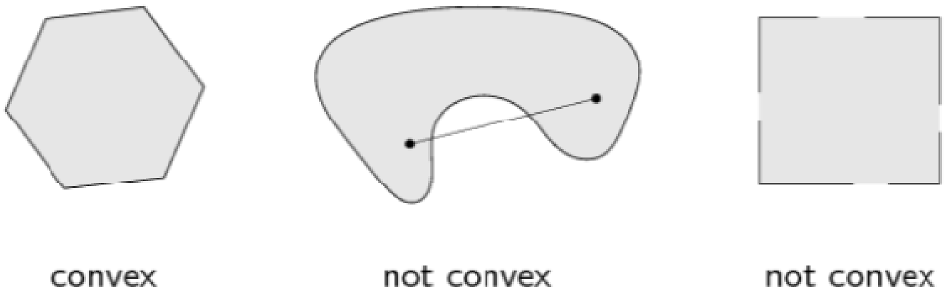
Also called convex

concave



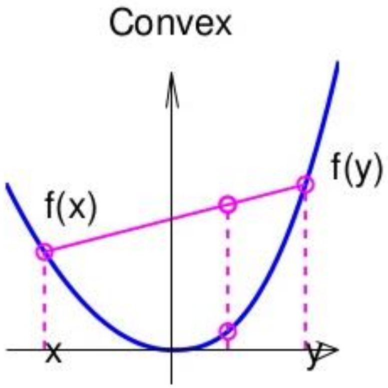
Convexity

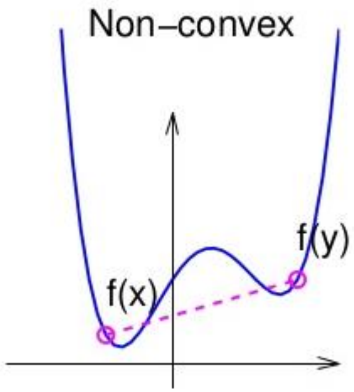
- Convex set
  - Imagine a set in R^2

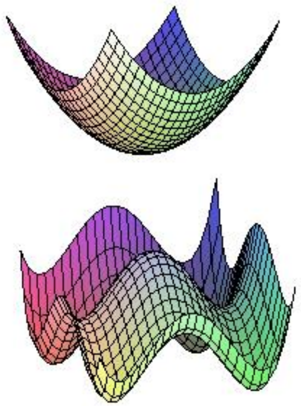


(Boyd, 2014, Convex Optimization)

- Convex function







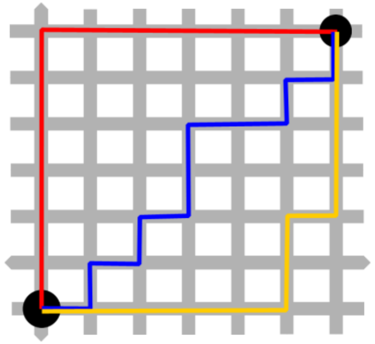
Norm

- A way to measure the length of a vector
  - Euclidean norm: $\| x \|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$

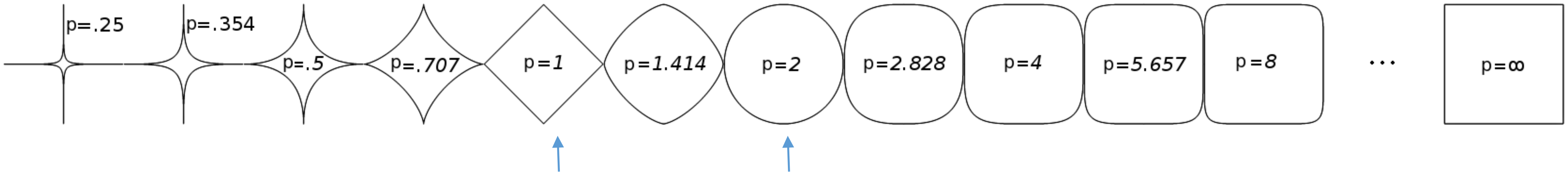
P-norm:

$$\| x \|_p = \left(\sum_1^n x_i^p \right)^{\frac{1}{p}}$$

- Other ways to measure the length of a vector
  - Taxicab norm: $\| x \|_1 = | x_1 | + | x_2 | + \ldots + | x_n |$



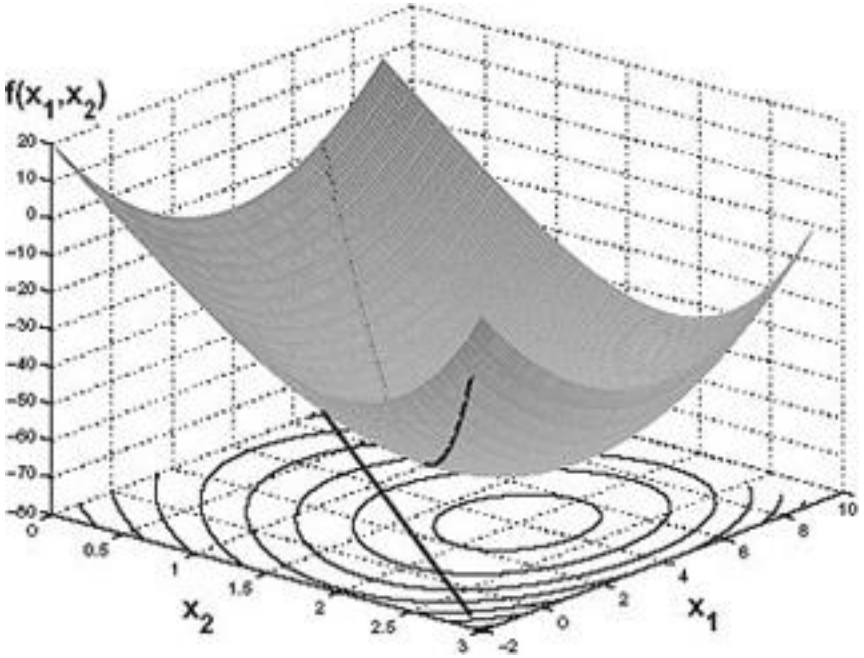
- Norm balls
  - Set of all the vectors emanated from the origin in Euclidean space R^2 with equal length (e.g. $\| x \|_p \leq 1$)



Convex Optimization Problem

- f : convex function C : convex set

$$\underset{x}{\text{Minimize}} \quad f(x) \quad \text{such that} \quad x \in C$$



(MIT open courseware, 2008, non-linear programming)

- Why do we care about convex optimization?
  - We know how to compute them
  - Instead of finding minimum of a function that looks like 
  - It's much easier if a function looks like  , where its local minimum is also its global minimum.

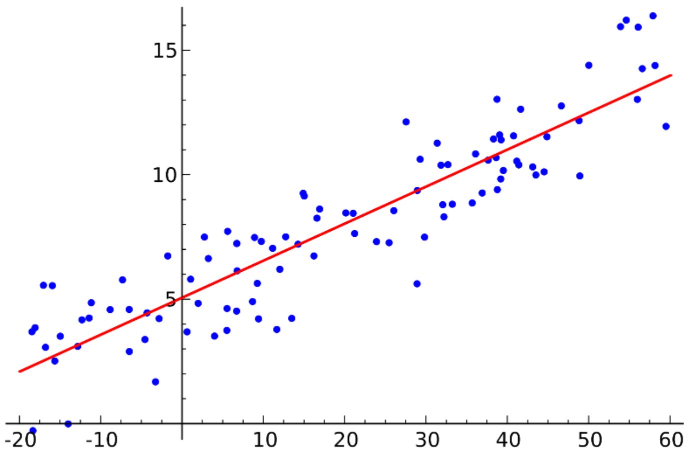
Sparsity

- What is sparsity?
  - Vector in R^n that has lots of zeros. (e.g. $\begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$)

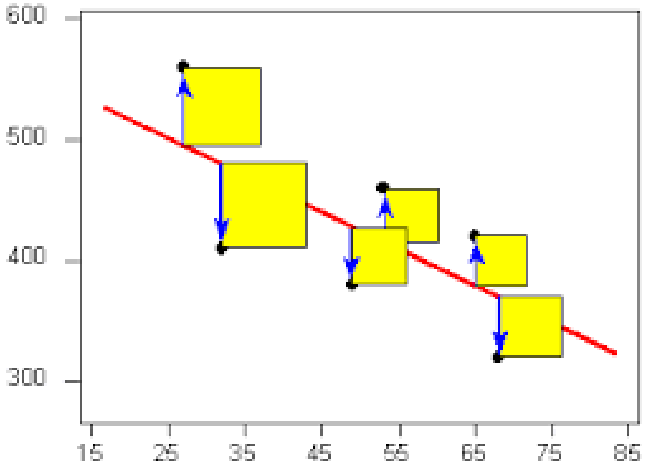
  - Useful in signal processing (e.g. images, sound waves, etc.)
  - Saves computational resource (Memories, Floating-point operations).

What kind of techniques in convex optimization give us a sparse solution?

Recall: Linear Regression



Common way to solve this: Ordinary Least Squares

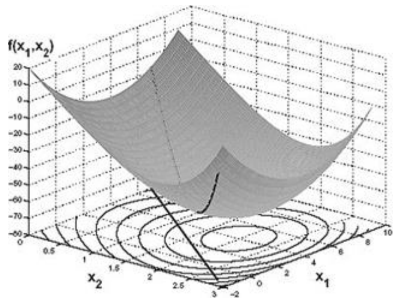


A convex optimization problem

Minimize $\sum_i \varepsilon_i^2$

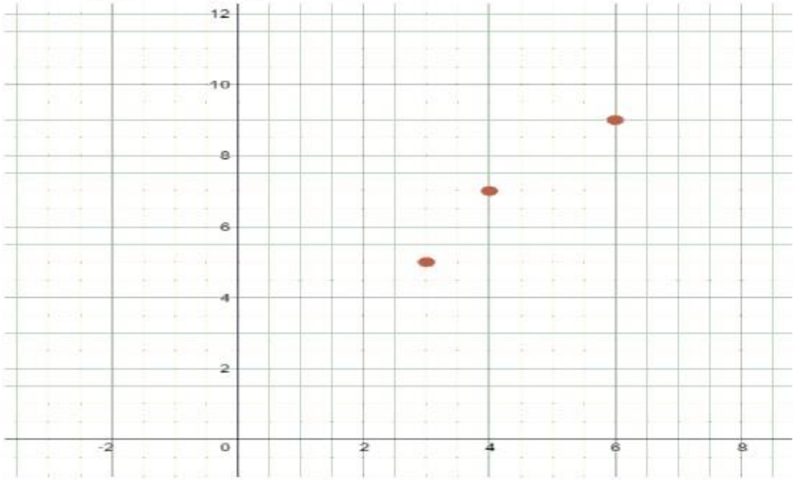
Minimize$_\theta$ $\| y - X\theta \|_2^2$

Least squares minimization

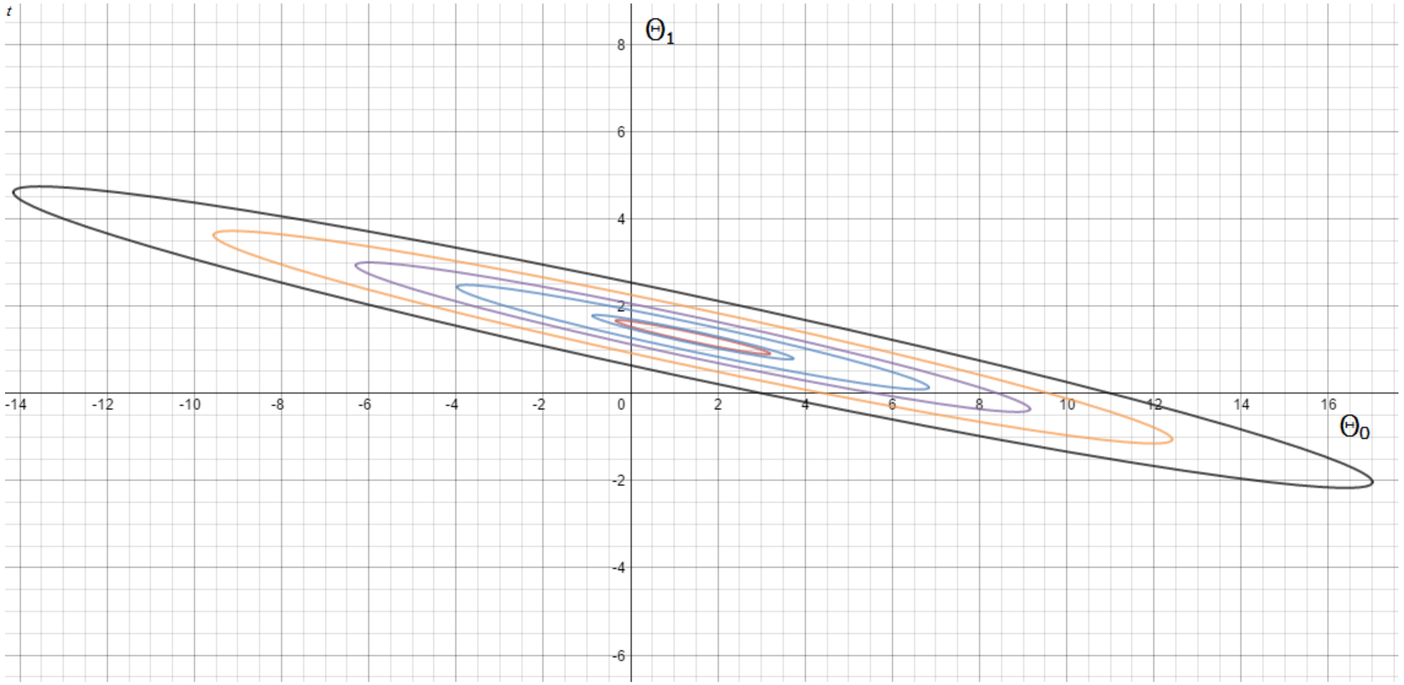


Rooms (X)	Price (Y, \$10,000)
3	5
4	7
6	9

The graph of the regression problem we have:



$$\min_{\theta \in R^2} \| y - X\theta \|_2^2 = \min_{\theta \in R^2} f(\theta_0, \theta_1)$$

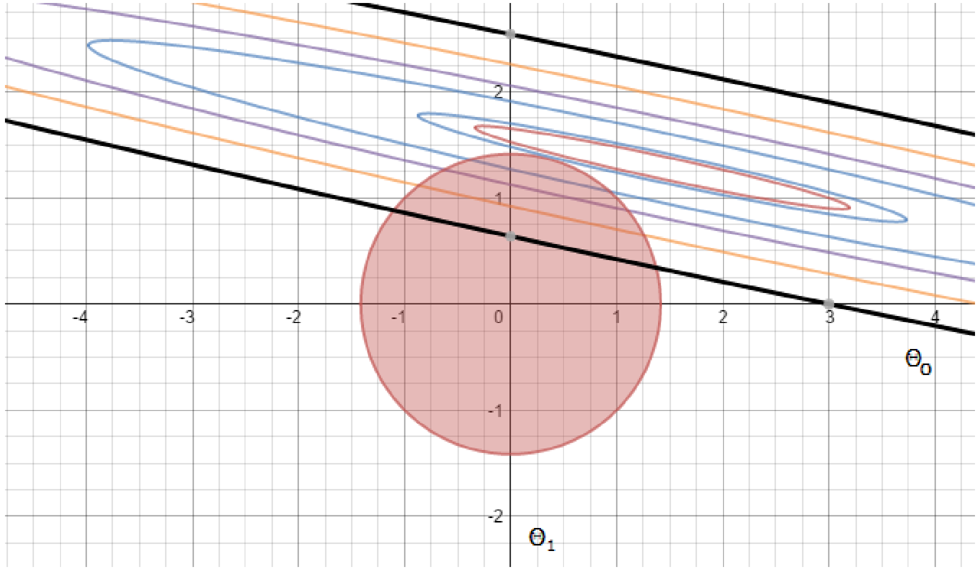


Ridge Regression and Lasso Regression

$f(\theta_0, \theta_1) = \| y - X\theta \|_2^2$

Ridge Regression

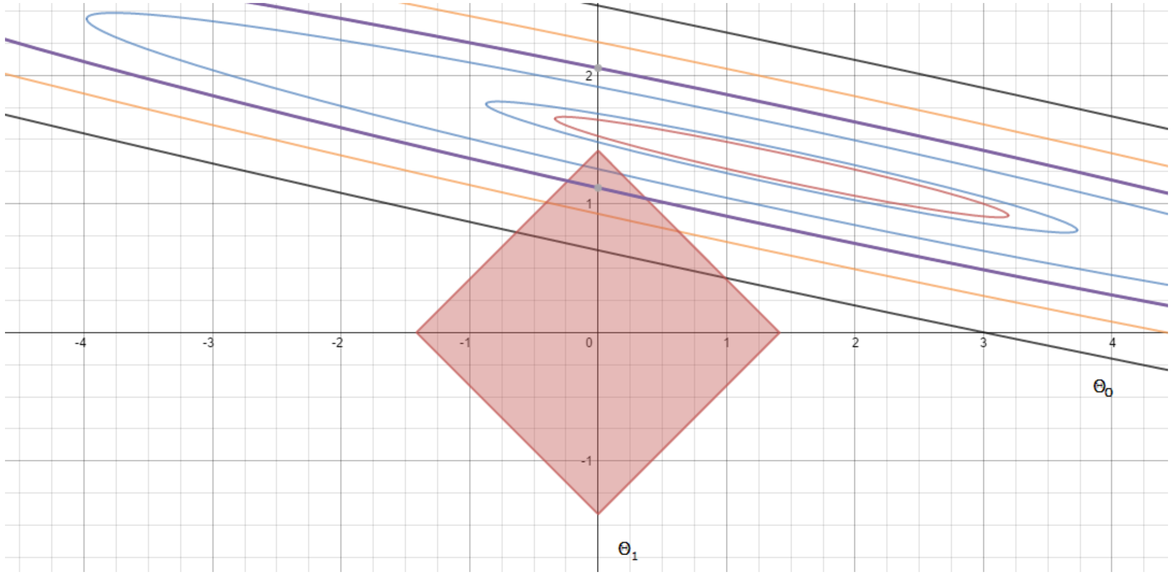
$$\min_{\theta \in R^2} f(\theta_0, \theta_1) s.t. \quad \| \theta \|_2 \leq t$$



The minimizer (θ_0, θ_1) does not lie on the coordinate axis

Lasso Regression

$$\min_{\theta \in R^2} \| y - X\theta \|_2^2 \; s.t. \; \| \theta \|_1 \leq t$$



The minimizer (θ_0, θ_1) lie on the coordinate axis;
the minimizer $(\theta_0, \theta_1) = (0, t)$

Note: I implemented a LASSO solver called Least Angle Regression

Dictionary Learning

Called dictionary

- Take images $y_1,\dots y_m$, where each $y_i \in R^n$. The goal is a make a matrix $D \in R^{n\times k}$

such that $y_1,\dots y_m$ can be "well approximate" by $D\alpha_i$ for some sparse

vector $\alpha_1,\dots,\alpha_m \in R^k$.

- The images could be $y_1 =$ 

, $y_2 =$ 
, $y_3 =$ 
, $y_4 =$ 

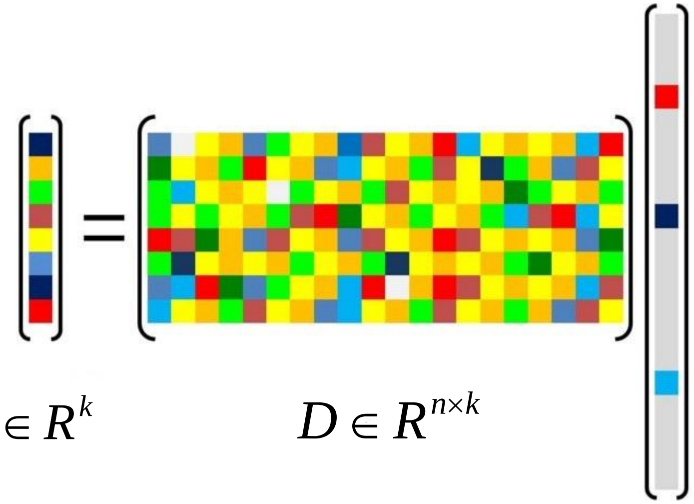
…and so on.

- If we input an image $y_{input} =$ 

, the algorithm, using the trained dictionaries, should tell us that it is a Jayhawk, not a  .

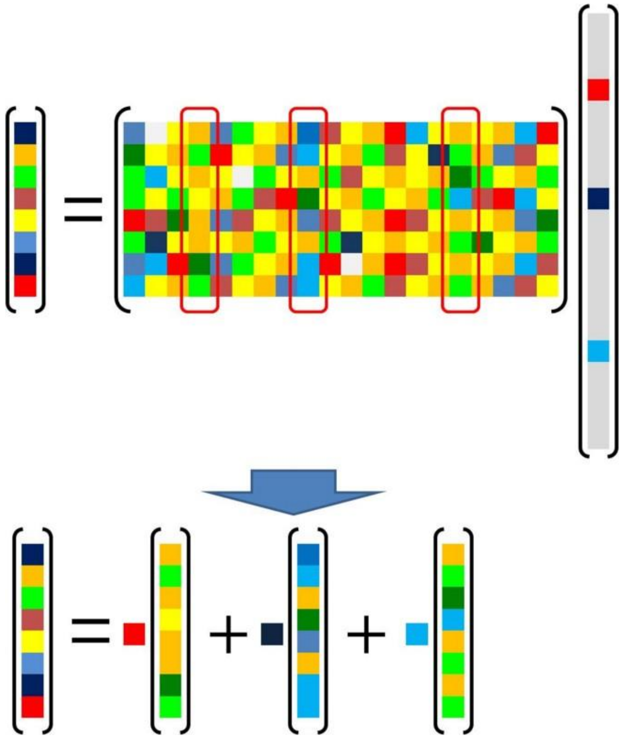
Dictionary Learning

- The matrix $D \in R^{n \times k}$ is called a "Dictionary".



$y \in R^k$ $D \in R^{n \times k}$ $\alpha \in R^n$

What makes a good dictionary?

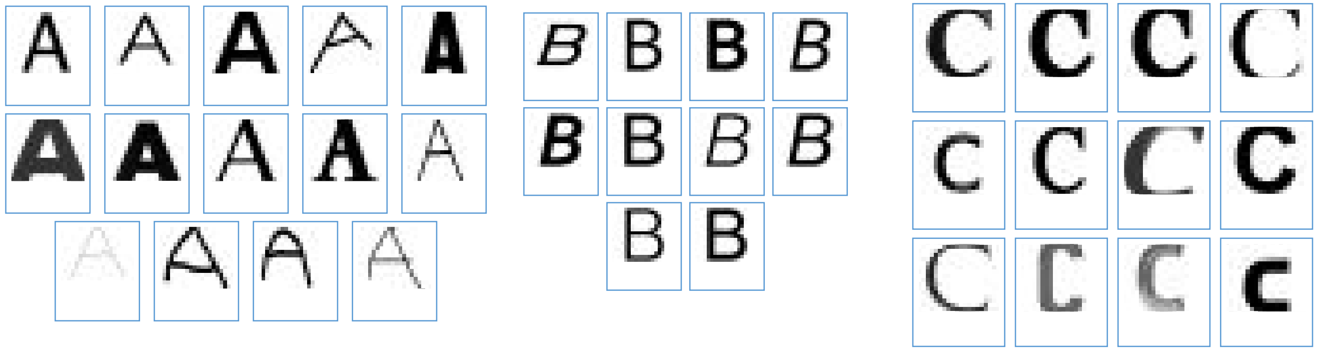


One that makes a good sparse representation of the signal.

(Sparse Representation and Compressed Sensing: Theory and Algorithms Yi Ma 1,2 Allen Yang 3 John Wright 1 CVPR Tutorial, June 20, 2009 1 Microsoft Research)

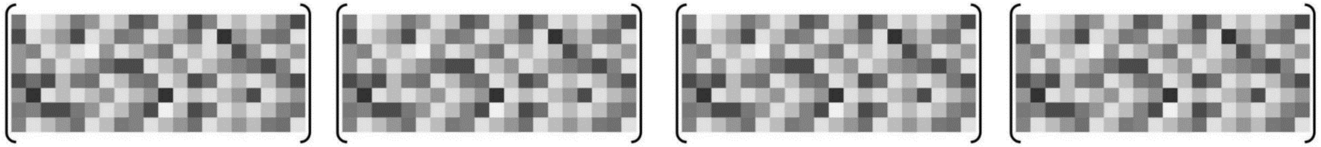
Implement Dictionary Learning

- Character recognition:
    - Given images letters A~Z



…

- Train a set of dictionaries using Gradient descent and LASSO



…

D_A D_B D_C D_D

To test the trained dictionary, input a character image, then use the trained dictionary to run LASSO and select the one the returns the least squared error.

Ex. Input image:

Reconstructions: