Class Notes

Dustin Leatherman

May 23, 2020

Outline

Review & Introduction (2020/03/31)

Review

Orthogonal: Vectors are orthogonal when the dot product = 0.

Basis

$$
\begin{aligned}
\underset{(n\times 1)}{\vec{y}} &= \underset{(n\times p)}{A}\underset{(p\times 1)}{\vec{x}} \\
&= B\vec{c} \\
&= \Sigma c_i \vec{b_i} \text{ (most } c_i = 0)
\end{aligned} \tag{1}
$$

A: Basis Matrix

Properties of a Good Basis

- not all are orthogonal
- Allows for a sparse vector to be used ad the constant vector $\vec{c}$

Identity Matrices are the *worst* basis because most coefficients are non-zero.

2-Sparse Vector

Why Separating Hyperplane Theorem & Subspace Segmentation Example (2020/04/07)

Why is Separating Hyper-plane Theorem true?

Math Background

Let $x = d - c, \; y = u - d$

Square of the l_2-norm is the inner product

$$\|x\|_2^2 = \langle x, x \rangle = x^T x$$

$$(d-c)^T (d-c) = \|d-c\|_2^2$$

Expansion of Vectors

$$\begin{aligned} & \|x + ty\|_2^2 \\ =& \langle x + ty, x + ty \rangle \end{aligned} \tag{12}$$

Sparse Representation & Problem P0 . P1 (2020/04/14)

Big Idea

Your Data is a vector $x \in R^N$ where all vectors are column vectors.
Each x is s-sparse i.e. each vector has at most $\underline{s}$ non-zero entries.
Let s = 5000. We don't know where the non-zero entries are located.

Let $\underset{(m \times N)}{A}$, $m < N$

$N = 100,000, \ m = 20,000$

Short + Wide Matrix

> *This is the opposite of the kinds of matrices seen in Linear Regression which are tall and skinny.*

What if we can design a matrix $A \in R^{m \times N}$ so that for each s-sparse $\vec{x} \in R^N$, you can store $\vec{y}$ instead? ($A\vec{x} = \vec{y}$)

Q: Is there a way to get back $\vec{x}$ from $\vec{y}$? We observe $\vec{y}$.

A: Yes!

Properties of A

- A cannot be the 0 matrix.

Sparse Representation pt 2 (2020/04/21)

Historical Perspective

Why is the visual system so powerful? Hypothesis is our brain uses sparse representation of Visual Data.
Let a picture $\vec{y} = c_1 \vec{b}_1 + ... + c_n \vec{b}_n$
so that most c_j are zero.
Sparse representation used to be called Sparse Coding.
Robust Facial Recognition uses Sparse Subspace Clustering.
Given 19 x 19 images, let $Y = [\vec{Y}_1 | ... | \vec{Y}_{45}]$, $\vec{y}_j \in R^{361}$
19 * 19 = 361
Given Y, solve for matrix C

$$Y = YC, \; diag(C) = \vec{0}$$

Since we don't want $Y_i = Y_i$, that is why the constraint $diag(C) = \vec{0}$ is introduced. It ensures that a group of vectors can be a linear combination of others.
Each column of C is sparse since we want all column vectors to be a linear combination of a smaller set of columns.

Sparse Representation Pt 3 (2020/04/28)

Expanding on RIP

Expanding upon RIP
Any S columns of the matrix A are nearly orthogonal to each other.

Expanding on IHT

Expanding upon the IHT Algorithm,
$\tau_x(\cdot)$ is an non-linear operator that outputs a sparse matrix. The operator is non-linear because it does not *change* the dimensions on the vector. i.e. $R^n \rightarrow R^n$. You will not be able to find a matrix that will return the same output as this operator.
$\tau_s(\vec{x}_1) = x_2$
Which means both $\vec{x}_1$ and $\vec{x}_2$ have an inner product.
The IHT algorithm is described below:

$$\vec{u}_n = \vec{x}_n + A^*(\vec{y} - A\vec{x}_n), \text{ where } \vec{y} = A\vec{x} \tag{28a}$$

$$= \vec{x}_n + (A^*A\vec{x} - A^*A\vec{x}_n) \tag{28b}$$

$$= (I - A^*A)\vec{x}_n + A^*A\vec{x} \tag{28c}$$

Gradient Descent (2020/05/05)

Method of Steepest Descent

Let $x \in R^3$, $y \in R^3$. these are column vectors in R^3

$$
\begin{aligned}
f(x) =& f(x_1, x_2, x_3) \\
f(y) =& f(y_1, y_2, y_3) \\
G(y) =& G(y_1, y_2, y_3)
\end{aligned} \tag{41}
$$

$\nabla f(x)$ is a gradient vector. The convention is that the gradient is a row vector.

$G(y) = f(y) - \nabla f(x) y$

$$
\begin{aligned}
\nabla f(x) \equiv& \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}\right) \\
\nabla f(x) y =& \frac{\partial f}{\partial x_1} y_1 + \frac{\partial f}{\partial x_2} y_2 + \frac{\partial f}{\partial x_3} y_3
\end{aligned} \tag{42}
$$

Lagrangian Multipliers (2020/05/12)

Prelude

Find MAX $x^2 + y^2$ subject to $x + y = 4$
Increase radius until it hits the slow of $x + y = 4$

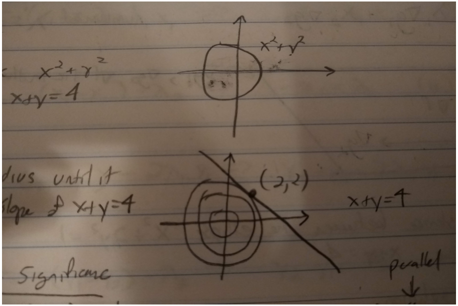

Figure: Prelude Drawgin

Geometric Significance

A $(x, y) = (2, 2)$ where MAX occurs: $\nabla f // \nabla g$