Topics in Applied Optimization: Lecture-2

Applications to Machine Learning, Vision, and Data Analytics

Pawan Kumar

CSTAR, IIIT-H

Optimization = Problem Solving

Consider the following temperature data of a city in a day:

Time	Rescaled Time	Temperature
9:00	0:00	15
11:00	2:00	30
14:00	4:00	34
16:00	6:00	33
18:00	8:00	21
22:00	12:00	18

Predict temperature at 14:30.

- Objective function:
- Constraints:
- Model:

Temperature Predictor: Modelling

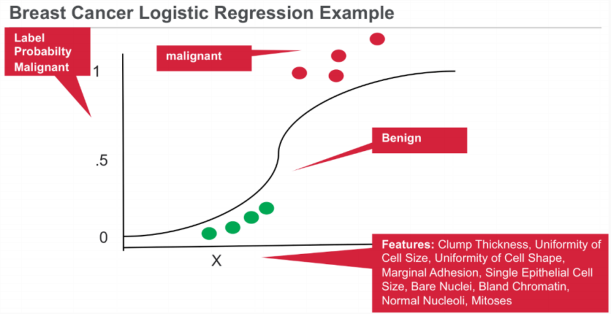
Plot the graph

Suggest a temprature predictor

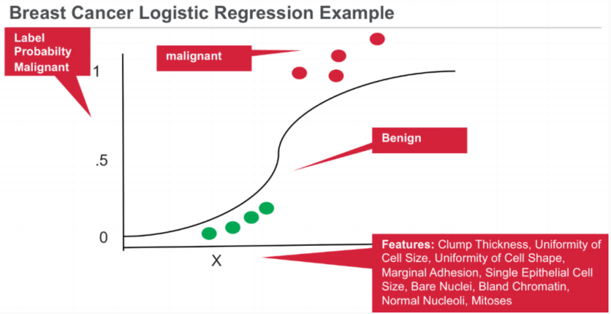
Solve the Optimization Problem: Least Squares

Optimization in Classification Problem: Logistic Regression

Optimization in Classification Problem: Logistic Regression



Optimization in Classification Problem: Logistic Regression

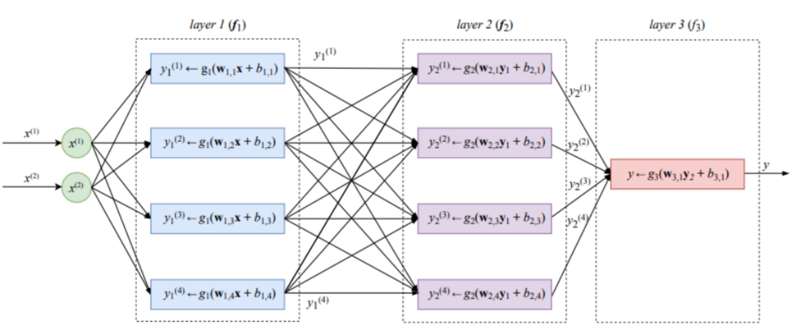


Objective (Loss) function:

$$J(\theta) = -1/m \sum \left[y^i \log(h_\theta(x^i) + (1 - y^i) \log(1 - h_\theta(x^i))) \right],$$

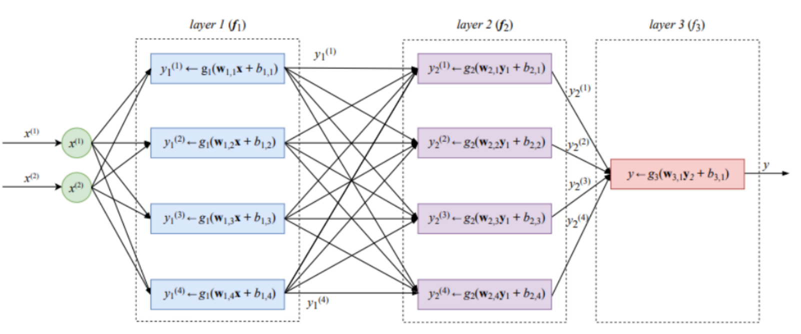
where, $h_\theta(x) = \dfrac{1}{1 + exp(-\theta^T x)}$

Optimization for Learning to Predict using Neural Networks



- NN is just another mathematical function $y = f_{NN}(x)$

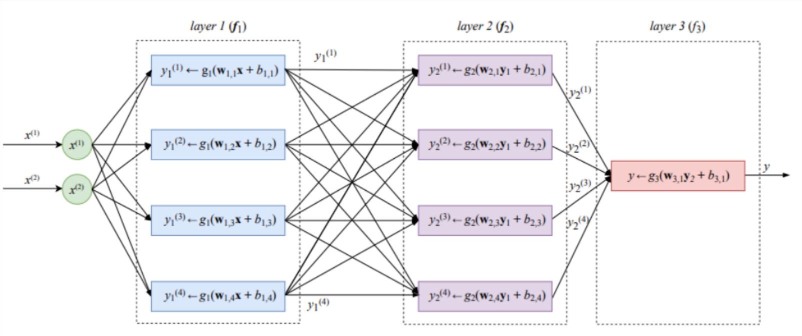
Optimization for Learning to Predict using Neural Networks



- NN is just another mathematical function $y = f_{NN}(x)$
- Here f_{NN} has nested form $f_{NN}(x) = f_3(f_2(f_1(x)))$, where f_1 and f_2 are vector functions of the form:

$$f_\ell = g_\ell(W_\ell z + b_\ell), \quad \ell = 1, 2.$$

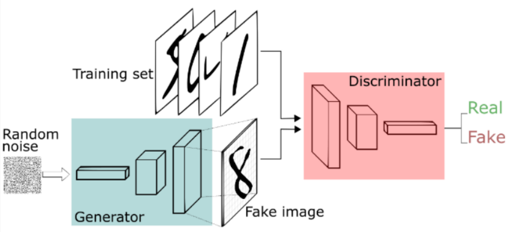
Optimization for Learning to Predict using Neural Networks



- NN is just another mathematical function $y = f_{NN}(x)$
- Here f_{NN} has nested form $f_{NN}(x) = f_3(f_2(f_1(x)))$, where f_1 and f_2 are vector functions of the form:

$$f_\ell = g_\ell(W_\ell z + b_\ell), \quad \ell = 1, 2.$$

Generative Models: GANs



$$
\begin{aligned}
E(G,D) &= \frac{1}{2}\mathbb{E}_{x\sim p_t}[1-D(x)] + \frac{1}{2}\mathbb{E}_{z\sim p_z}[D(G(z))] \\
&= \frac{1}{2}\left(\mathbb{E}_{x\sim p_t}[1-D(x)] + \mathbb{E}_{x\sim p_g}[D(x)]\right)
\end{aligned}
$$

- Objective function is:

$$\max_G(\min_D E(G, D))$$

Review of Linear Algebra

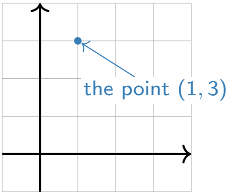
Section 3.1

Vectors

We have been drawing elements of $\mathbf{R}^n$ as points in the line, plane, space, etc. We can also draw them as arrows.

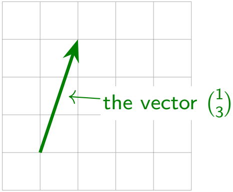
Definition

A **point** is an element of $\mathbf{R}^n$, drawn as a point (a dot).



A **vector** is an element of $\mathbf{R}^n$, drawn as an arrow. When we think of an element of $\mathbf{R}^n$ as a vector, we'll usually write it vectically, like a matrix with one column:

$$v = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

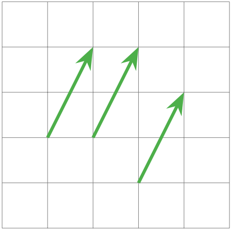


[interactive]

The difference is purely psychological: *points and vectors are just lists of numbers*.

So why make the distinction?

A vector need not start at the origin: *it can be located anywhere*! In other words, an arrow is determined by its length and its direction, not by its location.



These arrows all represent the vector $\begin{pmatrix}1\\2\end{pmatrix}$.

However, unless otherwise specified, we'll assume a vector starts at the origin.

Vector Algebra

Definition

- We can add two vectors together:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a+x \\ b+y \\ c+z \end{pmatrix}.$$

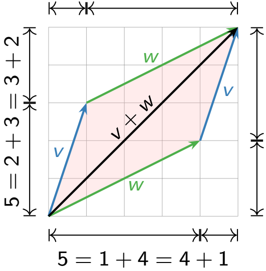
- We can multiply, or **scale**, a vector by a real number c:

$$c \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c \cdot x \\ c \cdot y \\ c \cdot z \end{pmatrix}.$$

  We call c a **scalar** to distinguish it from a vector. If v is a vector and c is a scalar, cv is called a **scalar multiple** of v.

(And likewise for vectors of length n.) For instance,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \quad \text{and} \quad -2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \\ -6 \end{pmatrix}.$$

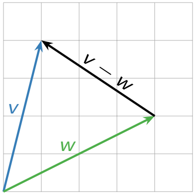


The parallelogram law for vector addition

Geometrically, the sum of two vectors v, w is obtained as follows: place the tail of w at the head of v. Then $v + w$ is the vector whose tail is the tail of v and whose head is the head of w. Doing this both ways creates a parallelogram. For example,

$$\begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}5\\5\end{pmatrix}.$$

Why? The width of $v + w$ is the sum of the widths, and likewise with the heights. [interactive]



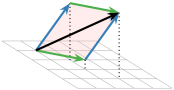
Vector subtraction

Geometrically, the difference of two vectors v, w is obtained as follows: place the tail of v and w at the same point. Then $v - w$ is the vector from the head of w to the head of v. For example,

$$\begin{pmatrix}1\\4\end{pmatrix} - \begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}-3\\2\end{pmatrix}.$$

Why? If you add $v - w$ to w, you get v. [interactive]

This works in higher dimensions too!

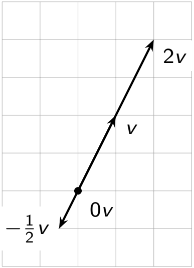


Scalar Multiplication: Geometry

Scalar multiples of a vector

These have the same *direction* but a different *length*.

Some multiples of v.



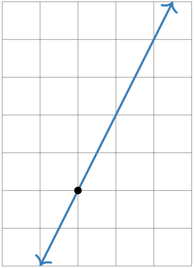
$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$2v = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$-\frac{1}{2}v = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$$

$$0v = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

All multiples of v.



[interactive]

So the scalar multiples of v form a *line*.

Linear Combinations

We can add and scalar multiply in the same equation:

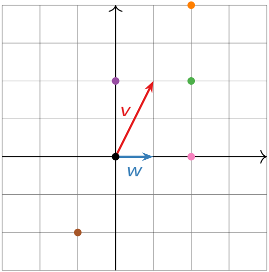
$$w = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p$$

where $c_1, c_2, \ldots, c_p$ are scalars, $v_1, v_2, \ldots, v_p$ are vectors in $\mathbf{R}^n$, and w is a vector in $\mathbf{R}^n$.

Definition

We call w a **linear combination** of the vectors $v_1, v_2, \ldots, v_p$. The scalars $c_1, c_2, \ldots, c_p$ are called the **weights** or **coefficients**.

Example



Let $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

What are some linear combinations of v and w?

- $v + w$
- $v - w$
- $2v + 0w$
- $2w$
- $-v$

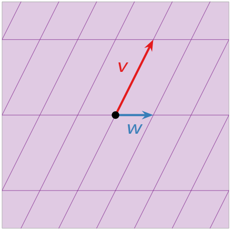
[interactive: 2 vectors] [interactive: 3 vectors]

Poll

Poll

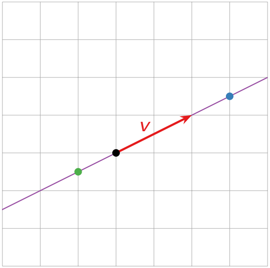
Is there any vector in $\mathbf{R}^2$ that is *not* a linear combination of v and w?

No: in fact, *every* vector in $\mathbf{R}^2$ is a combination of v and w.



(The purple lines are to help measure *how much* of v and w you need to get to a given point.)

More Examples

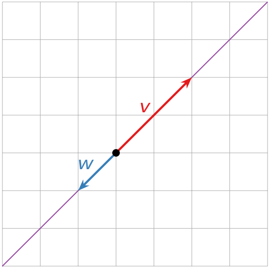


What are some linear combinations of $v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$?

- $\frac{3}{2}v$
- $-\frac{1}{2}v$
- ...

What are *all* linear combinations of v?

All vectors cv for c a real number. I.e., all *scalar multiples* of v. These form a *line*.



Question

What are all linear combinations of

$$v = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} -1 \\ -1 \end{pmatrix}?$$

Answer: The line which contains both vectors.

What's different about this example and the one on the poll? [interactive]

Section 3.2

Vector Equations and Spans

Solve the following system of linear equations:

$$\begin{aligned} x - y &= 8 \\ 2x - 2y &= 16 \\ 6x - y &= 3. \end{aligned}$$

We can write all three equations at once as vectors:

$$\begin{pmatrix} x - y \\ 2x - 2y \\ 6x - y \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}.$$

We can write this as a linear combination:

$$x \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} + y \begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}.$$

So we are asking:

Question: Is $\begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$ a linear combination of $\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix}$?

Systems of Linear Equations

Continued

$$\begin{array}{rcl} x - y &=& 8 \\ 2x - 2y &=& 16 \\ 6x - y &=& 3 \end{array} \quad \xrightarrow{\text{matrix form}} \quad \left(\begin{array}{rr|r} 1 & -1 & 8 \\ 2 & -2 & 16 \\ 6 & -1 & 3 \end{array}\right)$$

$$\xrightarrow{\text{row reduce}} \quad \left(\begin{array}{rr|r} 1 & 0 & -1 \\ 0 & 1 & -9 \\ 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{\text{solution}} \quad \begin{array}{rcl} x &=& -1 \\ y &=& -9 \end{array}$$

Conclusion:

$$-\begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} - 9\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$$

[interactive] $\longleftarrow$ (this is the picture of a *consistent* linear system)

What is the relationship between the vectors in the linear combination and the matrix form of the linear equation? They have the same columns!

Shortcut: You can go directly between augmented matrices and vector equations.

Vector Equations and Linear Equations

Summary

The **vector equation**

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = b,$$

where $v_1, v_2, \ldots, v_p, b$ are vectors in $\mathbf{R}^n$ and $x_1, x_2, \ldots, x_p$ are scalars, has the same solution set as the linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_p & b \\ | & | & & | & | \end{array}\right),$$

where the v_i's and b are the columns of the matrix.

So we now have (at least) *two* equivalent ways of thinking about linear systems of equations:

1. Augmented matrices.
2. Linear combinations of vectors (vector equations).

The last one is more geometric in nature.

Span

It is important to know what are *all* linear combinations of a set of vectors $v_1, v_2, \ldots, v_p$ in $\mathbf{R}^n$: it's exactly the collection of all b in $\mathbf{R}^n$ such that the vector equation (in the unknowns $x_1, x_2, \ldots, x_p$)

$$x_1v_1 + x_2v_2 + \cdots + x_pv_p = b$$

has a solution (i.e., is consistent).

Definition

Let $v_1, v_2, \ldots, v_p$ be vectors in $\mathbf{R}^n$. The **span** of $v_1, v_2, \ldots, v_p$ is the collection of all linear combinations of $v_1, v_2, \ldots, v_p$, and is denoted $\mathrm{Span}\{v_1, v_2, \ldots, v_p\}$. In symbols:

“the set of” “such that”

$$\boxed{\mathrm{Span}\{v_1, v_2, \ldots, v_p\} = \big\{ x_1v_1 + x_2v_2 + \cdots + x_pv_p \;\big|\; x_1, x_2, \ldots, x_p \text{ in } \mathbf{R} \big\}.}$$

Synonyms: $\mathrm{Span}\{v_1, v_2, \ldots, v_p\}$ is the subset **spanned by** or **generated by** $v_1, v_2, \ldots, v_p$.

This is the first of several definitions in this class that you simply **must learn**. I will give you other ways to think about Span, and ways to draw pictures, but *this is the definition*. Having a vague idea what Span means will not help you solve any exam problems!

Now we have several equivalent ways of making the same statement:

1. A vector b is in the span of $v_1, v_2, \ldots, v_p$.
2. The vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_p v_p = b$$

   has a solution.
3. The linear system with augmented matrix

$$\left(\begin{array}{cccc|c} | & | & & | & | \\ v_1 & v_2 & \cdots & v_p & b \\ | & | & & | & | \end{array}\right)$$

   is consistent.

[interactive example] $\longleftarrow$ (this is the picture of an *inconsistent* linear system)

Note: **equivalent** means that, for any given list of vectors $v_1, v_2, \ldots, v_p, b$, *either* all three statements are true, *or* all three statements are false.