Linear algebra: Review

Equation of a 2D line

Implicit equations

slope ← intercept_y

1. $y = mx + c$ $\qquad x=0 \qquad$ 2 parameters

2. $ax + by + c = 0$ $\qquad$ 3 parameters

Line

$$L(a,b,c) = \{(x,y) : ax + by + c = 0,\ x \in \mathbb{R},\ y \in \mathbb{R}\}$$

$$C(x_0, y_0, r) = \{(x,y) : (x - x_0)^2 + (y - y_0)^2 = r^2,\ x, y \in \mathbb{R}\}$$

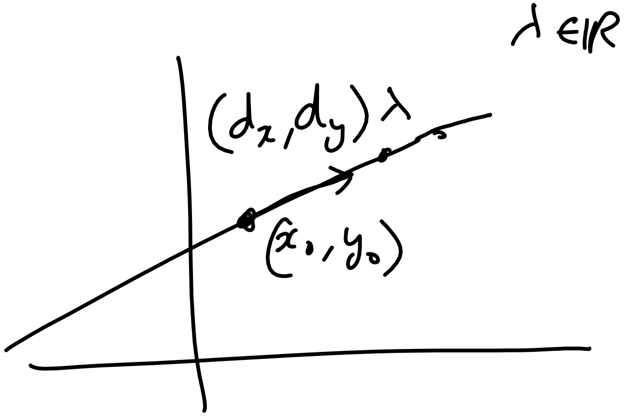
Circle

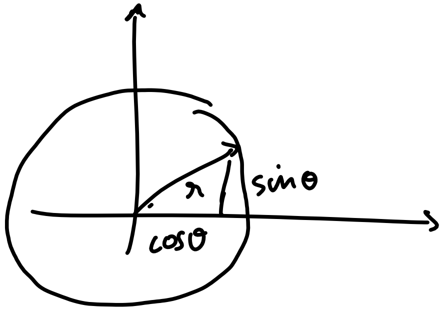
Implicit form

$$L(d_x, d_y, x_0, y_0) = \{(\lambda d_x + x_0, \lambda d_y + y_0) \ \forall \lambda \in \mathbb{R}\}$$

$$C(x_0, y_0, r) = \{(r\cos\theta + x_0, r\sin\theta + y_0) \ \forall \theta \in [0, 2\pi)\}$$

Paramteric form of 2D Line

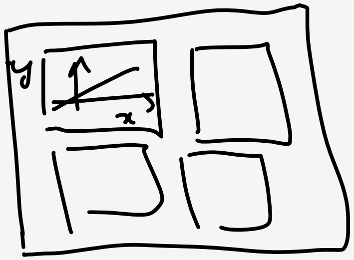




Matplotlib

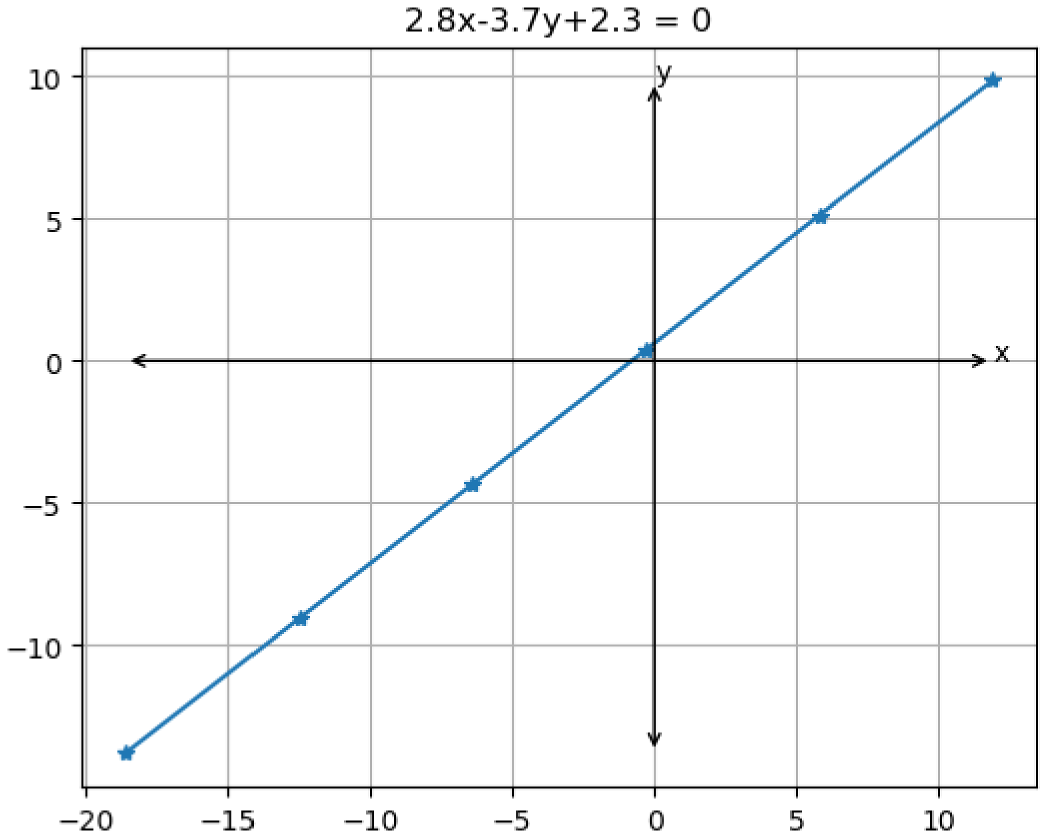
In [13]:

```
# Plot a line ax + by + c = 0
# a, b, c = 2.5, -1, -5 # pick numbers by hand

# pick a, b, c at random
import random
scale = 10
a, b, c = [scale*(random.random()-0.5) for _ in range(3)] # random numb

# Generate some sample points on a line
x, y = points_on_line(a, b, c, scale=scale)

# Plot the points
fig, ax = plt.subplots()
stylizeax(ax, (min(x), max(x), min(y), max(y)))
ax.plot(x, y, '*-') # the line
ax.set_title(f'{a:.1f}x{b:+.1f}y{c:+.1f} = 0') # print the equation
```



Out[13]:

```
Text(0.5, 1.0, '2.8x-3.7y+2.3 = 0')
```



2.8x-3.7y+2.3 = 0
y
x
10
5
0
−5
−10
−20
−15
−10
−5
0
5
10

Vectors

$\underbrace{\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{\text{geometric vector}}$

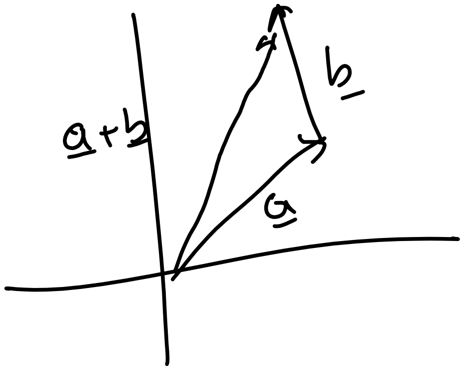
$\underline{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ ← size

Real

$\underline{a} \in \mathbb{R}^2$

$$\underline{a} + \underline{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1+2 \\ 2+3 \end{bmatrix}$$



Dot product $\quad \underline{a} \cdot \underline{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 1 \cdot 3 + 2 \cdot 4$

$\underline{a} \in \mathbb{C}^n \qquad \mathbb{R}^n$

$$\underline{a} \cdot \underline{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$$

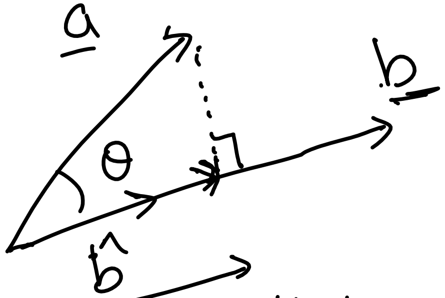
$\underbrace{\underline{a} \cdot \underline{b}}_{\text{scalar}} \in \mathbb{C}$

$\underline{a} \cdot \underline{b} \in \mathbb{R}$ scalar

Geometric interpretation of Dot product

$\underline{a} \cdot \underline{b}$

$\underline{b} = \begin{cases} \rightarrow \text{magnitude } \|\underline{b}\| \text{ norm} \\ \rightarrow \text{direction } \hat{b} = \dfrac{\underline{b}}{\|\underline{b}\|} \end{cases}$



$\underline{a} \cdot \hat{b} = \|\underline{a}\| \cos\theta$

$\underline{a} \cdot \hat{b}$

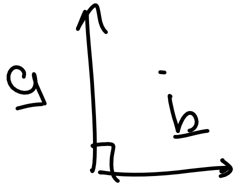
$\underline{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \|b\| = \sqrt{1^2 + 2^2}$

$\|b\| = \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}$

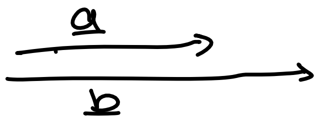
n-D vector

$\cos(0) = 1$

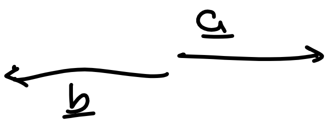
$\cos(90^\circ) = 0$



$\underline{a} \cdot \underline{b} = 0$



$\underline{a} \cdot \underline{b} = \|\underline{a}\| \|\underline{b}\|$



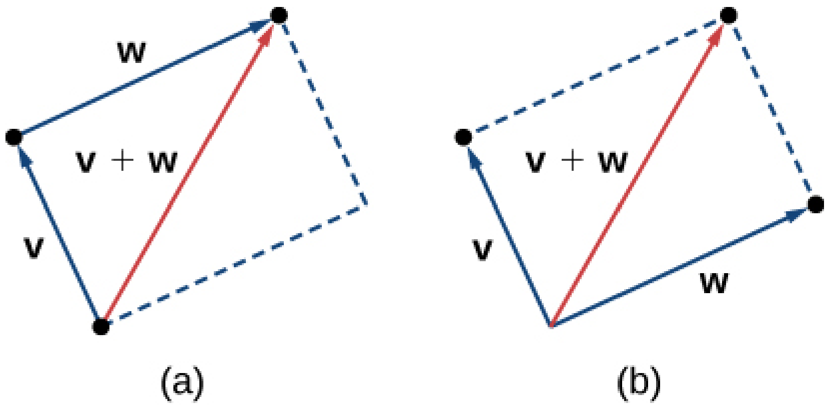
$\underline{a} \cdot \underline{b} = -\|\underline{a}\| \|\underline{b}\|$

Vector addition

Vector addition is element-wise addition

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{bmatrix}$$

Geometrically the resulting vector can be obtained by triangle law or the parallelogram law.



Reference: [1]

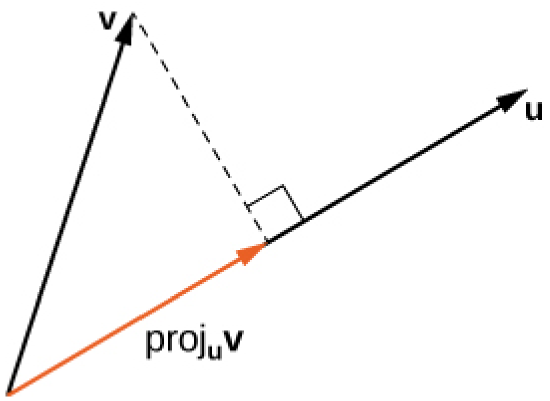
Dot product of vectors

Dot product of two vectors is a scalar given by sum of element-wise product.

$$\mathbf{v} \cdot \mathbf{u} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = v_1 u_1 + v_2 u_2 + \cdots + v_n u_n$$

Geometrically, dot product is closely related to the projection. Projection of vector $\mathbf{v}$ on $\mathbf{u}$ is the dot product of $\mathbf{v}$ with the direction of $\mathbf{u}$

$$\text{proj}_{\mathbf{u}}\mathbf{v} = \mathbf{v} \cdot \hat{\mathbf{u}}$$



Dot product of vector with itself gives the square of the magnitude $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$.

Reference: [2]

Matrices

Transpose of a Matrix

Tranpose of a column vector

Matrix-vector product

Matrix-matrix product

Identity matrix

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Square matrix

A square matrix is a matrix with number of rows equal to the number of columns.

Inverse of a square matrix

A matrix $\mathbf{V}^{-1}$ is called the inverse of a square matrix $\mathbf{V}$ if $\mathbf{V}^{-1}\mathbf{V} = \mathbf{V}^{-1} = \mathbf{I}_n$. The inverse of a square matrix exists only when it is singular i.e the determinant of the matrix is non-zero $\det(\mathbf{V}) \neq 0$.

Using vectors for 2D line notation

$\underbrace{ax+by}$ $+ c = 0$

pts $\quad \underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\underline{w} = \begin{bmatrix} a \\ b \end{bmatrix}$

weights

$\underline{w} \cdot \underline{x} = ax+by$

$\underline{w} \cdot \underline{x} + c = 0$

for the same line

we have ∞ many equations

$5x + 2y - 10 = 0$

$10x + \cancel{4}y - 20 = 0$

$ax + by + c = 0$

for any $\alpha \neq 0$

$\alpha a x + \alpha b y + \alpha c = 0$

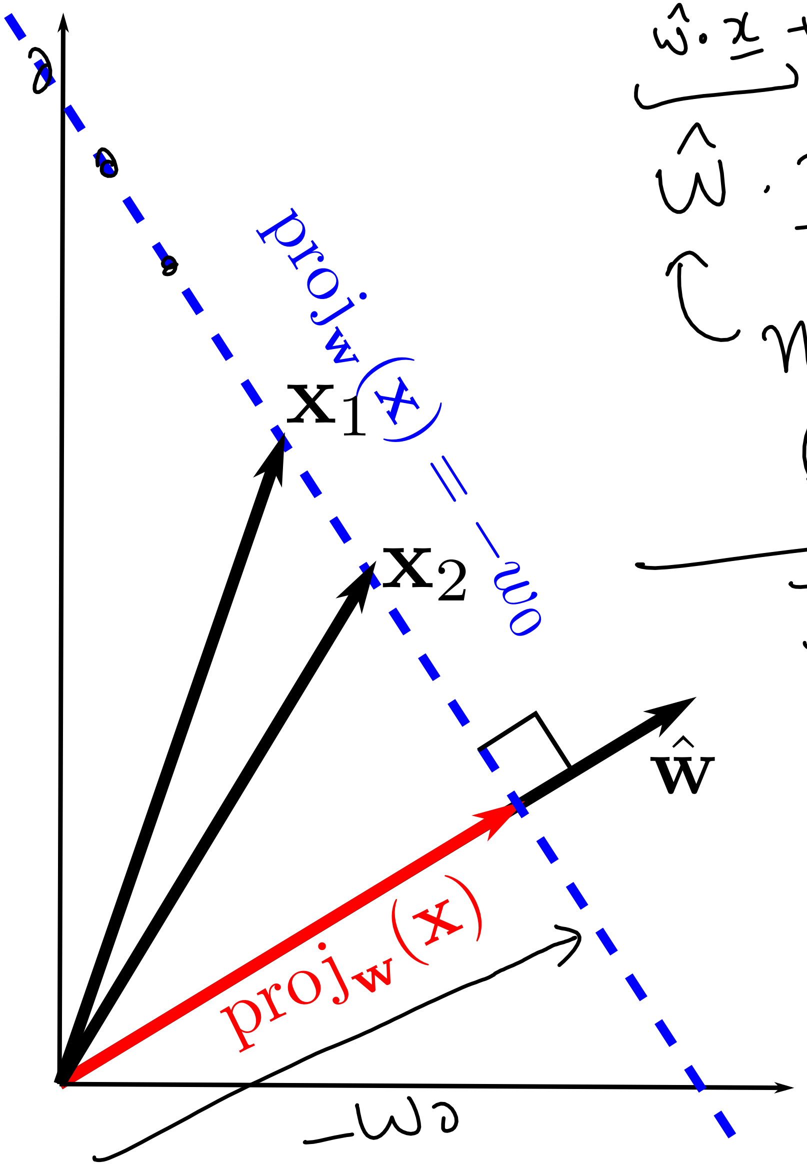
$\alpha = \frac{1}{\|w\|}$ —

Geometric interpretaion , $\frac{w \cdot x}{\|w\|} + \frac{c}{\|w\|} = 0$ $\hat{w} \in \mathbb{R}^2$

$\underbrace{\hat{w} \cdot \underline{x}} + w_0 = 0$ $w_0 = \frac{c}{\|w\|}$ $\underline{x} \in \mathbb{R}^2$

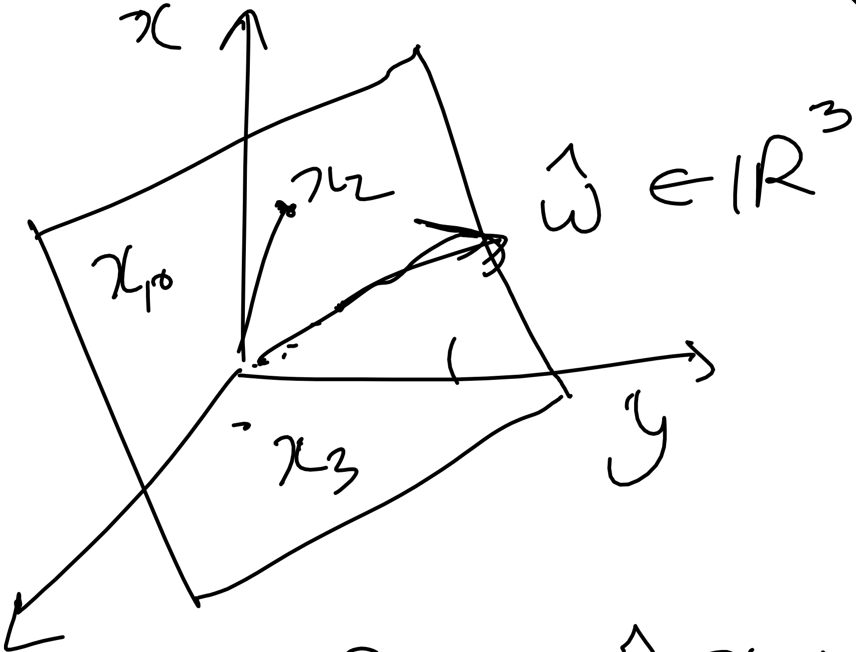
$\hat{w} \cdot \underline{x} = -w_0$

$\hat{}$ normal to the line



Implicit eq of line in 2D extends to implicit eq of plane in 3D

$\hat{w} \cdot \underline{x} = -w_0$



$\hat{w} \in \mathbb{R}^3$

$\underline{x} \in \mathbb{R}^3$

Line

plane

$L(\hat{w}, w_0) = \{ \underline{x} : \hat{w} \cdot \underline{x} + w_0 = 0 \, ; \, \underline{x} \in \mathbb{R}^2 \}$

$P(\hat{w}, w_0) = \{ \quad '' \quad \underline{x} \in \mathbb{R}^3 \}$

$HP(\hat{w}, w_0) \quad \{ \quad '' \quad \underline{x} \in \mathbb{R}^n \}$

Hyper plane

In [14]:

```python
import numpy as np # a vector algebra library

a = np.array([0, 1, 2, 3]) # a vector
print("a=", a)
b = np.array([4, 5, 6, 7]) # another vector
print("b=", b)
C = np.array([[0, 1, 2, 3],
              [4, 5, 6, 7]]) # A matrix
print("C=", C)
D = np.zeros((2, 4)) # a 2x4 matrix of zeros
print("D=", D)
E = np.random.rand(2,5) # Random 2x5 matrix of numbers between 0 and 1
print("E=", E)
```

```
a= [0 1 2 3]
b= [4 5 6 7]
C= [[0 1 2 3]
 [4 5 6 7]]
D= [[0. 0. 0. 0.]
 [0. 0. 0. 0.]]
E= [[0.24267745 0.34908614 0.31547851 0.15059988 0.17537179]
 [0.60868919 0.31716426 0.10530595 0.53841394 0.49799488]]
```

In [15]:

```python
print("a*0.1 = ", a * 0.1) # element-wise multiplication
print("C*0.2 = ", C * 0.2) # element-wise multiplication
print("a*b = ", a * b)   # element-wise multiplication (Note: different
print("a*b*0.2 = ", a * b * 0.2)   # element-wise multiplication
print("C @ a = ", C @ a)   # matrix-vector product
print("C.T = ", C.T)     # matrix transpose
print("C.T @ D = ", C.T @ D) # matrix-matrix product
print("a * C = ", a * C)   # so called broadcasting; numpy specific
```

```
a*0.1 =  [0.  0.1 0.2 0.3]
C*0.2 =  [[0.  0.2 0.4 0.6]
 [0.8 1.  1.2 1.4]]
a*b =  [ 0  5 12 21]
a*b*0.2 =  [0.  1.  2.4 4.2]
C @ a =  [14 38]
C.T =  [[0 4]
 [1 5]
 [2 6]
 [3 7]]
C.T @ D =  [[0. 0. 0. 0.]
 [0. 0. 0. 0.]
 [0. 0. 0. 0.]
 [0. 0. 0. 0.]]
a * C =  [[ 0  1  4  9]
 [ 0  5 12 21]]
```

$a = [0, 1, 2, 3]$

$C = [[\quad], [\quad]]$

$$a * c = \{ [\quad], [\quad] \}$$

Numpy: General Broadcasting Rules

When operating on two arrays, NumPy compares their shapes element-wise. It starts with the trailing (i.e. rightmost) dimension and works its way left. Two dimensions are compatible when

1. they are equal, or
2. one of them is 1.

Otherwise a ValueError is raised

Ref: https://numpy.org/doc/stable/user/basics.broadcasting.html

In the following example, both the A and B arrays have axes with length one that are expanded to a larger size during the broadcast operation:

```
A      (4d array):  8 x 1 x 6 x 1   → repeat  →   8 x 7 x 6 x 5
B      (3d array):      7 x 1 x 5               →  8 x 7 x 6 x 5
Result (4d array):  8 x 7 x 6 x 5
```

In [16]:
```
A = np.random.rand(8, 1, 6, 1)
B = np.random.rand(7, 1, 5)
(A * B).shape # Returns the shape of the multi dimensional array
```

Out[16]:
```
(8, 7, 6, 5)
```

Here are some more examples:

```
A      (2d array):  5 x 4                 A + B
B      (1d array):  1 x 1
Result (2d array):  5 x 4

A      (2d array):  5 x 4
B      (1d array):      4
Result (2d array):  5 x 4

A      (3d array):  15 x 3 x 5
B      (3d array):  15 x 1 x 5
Result (3d array):

A      (3d array):  15 x 3 x 5
B      (2d array):       3 x 5
Result (3d array):  15 x 3 x 5

A      (3d array):  15 x 3 x 5
B      (2d array):       3 x 1
Result (3d array):  15 x 3 x 5
```

Linear regression: review

Let's take the simple linear regression example from STS332 textbook (uploaded on brightspace;page 300; Table 6-1).

"As an illustration, consider the data in Table 6-1. In this table, y is the salt concentration (milligrams/liter) found in surface streams in a particular watershed and x is the percentage of the watershed area consisting of paved roads."

In [19]:

```
%%writefile saltconcentration.tsv
#Observation	SaltConcentration	RoadwayArea
1	3.8	0.19
2	5.9	0.15
3	14.1	0.57
4	10.4	0.4
5	14.6	0.7
6	14.5	0.67
7	15.1	0.63
8	11.9	0.47
9	15.5	0.75
10	9.3	0.6
11	15.6	0.78
12	20.8	0.81
13	14.6	0.78
14	16.6	0.69
15	25.6	1.3
16	20.9	1.05
17	29.9	1.52
18	19.6	1.06
19	31.3	1.74
20	32.7	1.62
```

```
Writing saltconcentration.tsv
```

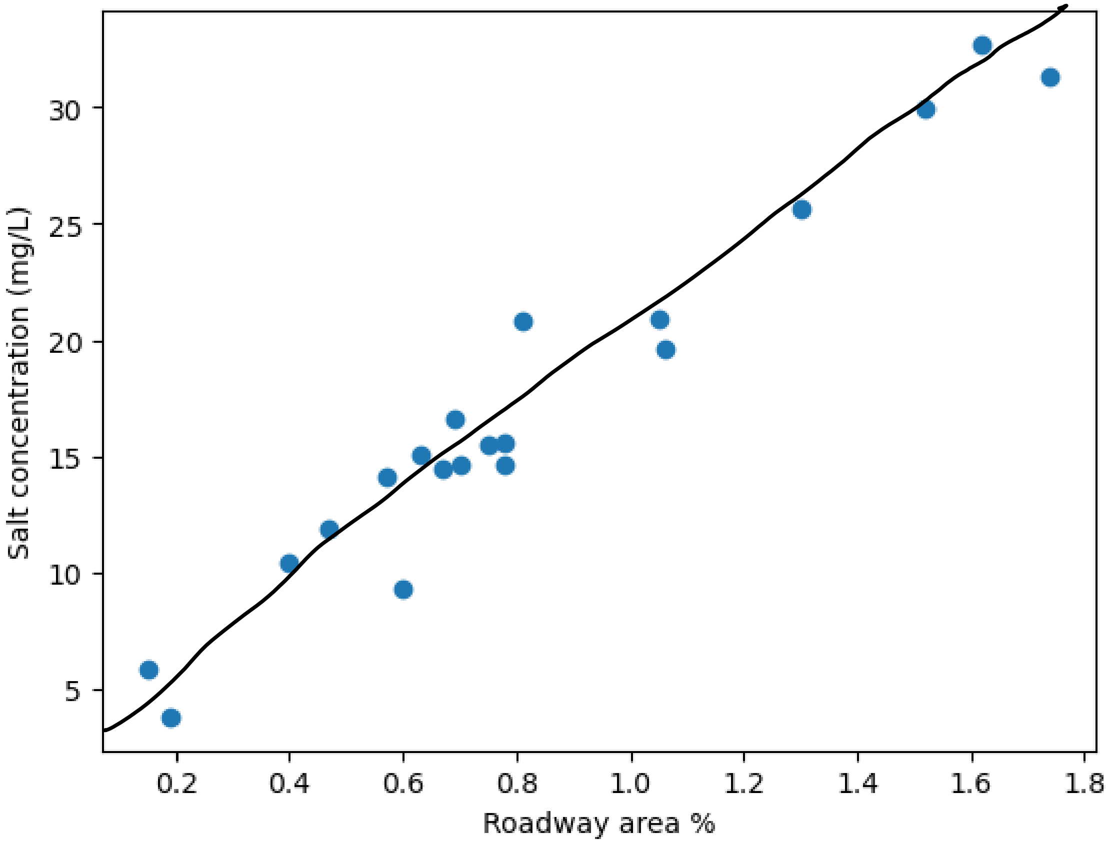
In [20]:
```
# numpy can import text files separated by seprator like tab or comma
salt_concentration_data = np.loadtxt("saltconcentration.tsv")
salt_concentration_data
```

Out[20]:
```
array([[ 1.  ,  3.8 ,  0.19],
       [ 2.  ,  5.9 ,  0.15],
       [ 3.  , 14.1 ,  0.57],
       [ 4.  , 10.4 ,  0.4 ],
       [ 5.  , 14.6 ,  0.7 ],
       [ 6.  , 14.5 ,  0.67],
       [ 7.  , 15.1 ,  0.63],
       [ 8.  , 11.9 ,  0.47],
       [ 9.  , 15.5 ,  0.75],
       [10.  ,  9.3 ,  0.6 ],
       [11.  , 15.6 ,  0.78],
       [12.  , 20.8 ,  0.81],
       [13.  , 14.6 ,  0.78],
       [14.  , 16.6 ,  0.69],
       [15.  , 25.6 ,  1.3 ],
       [16.  , 20.9 ,  1.05],
       [17.  , 29.9 ,  1.52],
       [18.  , 19.6 ,  1.06],
       [19.  , 31.3 ,  1.74],
       [20.  , 32.7 ,  1.62]])
```

x y

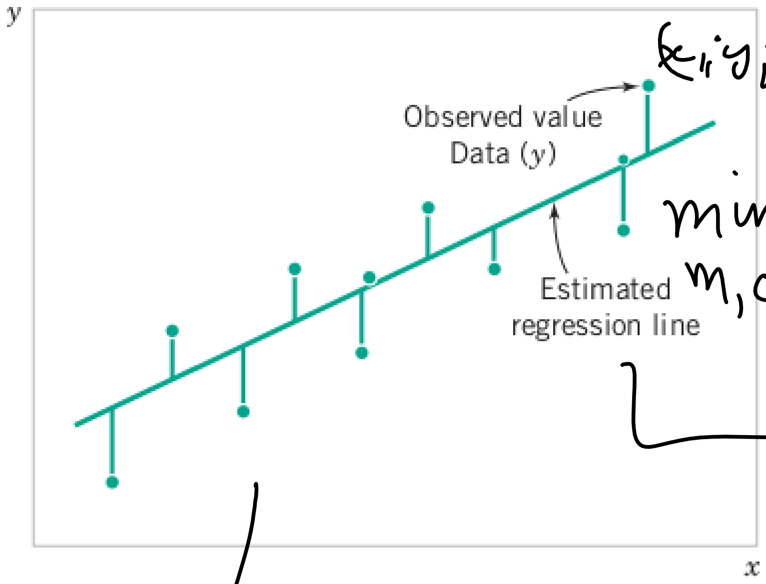
In [21]:
```python
# Plot the points
fig, ax = plt.subplots()
# Scatter plot using matplotlib
ax.scatter(salt_concentration_data[:, 2], salt_concentration_data[:, 1]
ax.set_xlabel(r"Roadway area %")
ax.set_ylabel(r"Salt concentration (mg/L)")
```

Out[21]:
```
Text(0, 0.5, 'Salt concentration (mg/L)')
```



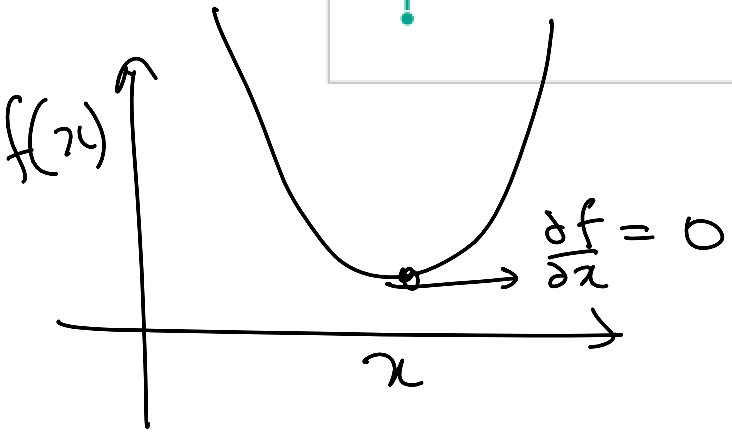
Least squares regression

→ "best fit" line to the data points



$$y_i = m x_i + c$$

$$\min_{m,c} \sum_{i=1}^{n} \Big(\underbrace{y_i - (m x_i + c)}_{\text{error}}\Big)^2$$



$$e_i = e(x_i, y_i) = y_i - (mx_i + c)$$

$$\underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$m^*, c^* = \arg\min_{m,c} \underbrace{\sum_{i=1}^{n} e_i^2}$$

$$\|\underline{e}\| = \sqrt{e_1^2 + e_2^2 + \dots + e_n^2}$$

$$= \arg\min_{m,c} \|\underline{e}\|^2$$

Vectorization of Least square regression

$$\|\underline{e}\|^2 = \underline{e} \cdot \underline{e}$$

$$\underline{e} \cdot \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$= e_1^2 + e_2^2 + \dots e_n^2$$

$$\underline{e} = \begin{bmatrix} y_1 - (mx_1 + c) \\ y_2 - (mx_2 + c) \\ \vdots \\ y_n - (mx_n + c) \end{bmatrix}$$

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad m\underline{x} = \begin{bmatrix} m x_1 \\ m x_2 \\ \vdots \\ m x_3 \end{bmatrix}$$

$$\underbrace{\underline{e} = \underline{y} - (m\underline{x} + c\underline{\mathbb{1}}_n)}_{\text{vectorization}} \qquad \mathbb{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$c = \begin{bmatrix} c \\ c \\ \vdots \\ c \end{bmatrix}$$

Matrices $\quad \underline{m} = \begin{bmatrix} m \\ c \end{bmatrix} \in \mathbb{R}^2$

Matrices

$$\underline{V} = m\left[\underline{v}_1, \underline{v}_2, \ldots \underline{v}_n \right] \qquad v_i \in \mathbb{R}^m$$

$$\underbrace{\qquad\qquad}_{n}$$

$\underline{V}$ $m \times n$ matrix $\qquad \underline{V} \in \mathbb{R}^{m \times n}$

Matrix transpose

$$\underline{V} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \rightarrow \underline{V}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$$\underline{V} \in \mathbb{R}^{m \times n} \rightarrow \underline{V}^T \in \mathbb{R}^{n \times m}$$

$$\underline{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^{m \times 1} \rightarrow \underline{u}^T \in \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \in \mathbb{R}^{1 \times m}$$

Matrix-vetor product

$$\underline{V} = \begin{bmatrix} \underline{v}_1^T \\ \underline{v}_2^T \\ \vdots \\ v_m^T \end{bmatrix}$$

$$\underline{V}\,\underline{u} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix} \underline{u} = \begin{bmatrix} \underline{v}_1^T \underline{u} \\ v_2^T \underline{u} \\ \vdots \\ v_m^T \underline{u} \end{bmatrix}$$

$\underline{v}_i^T \in \mathbb{R}^{1 \times n}$

$\underline{u} \in \mathbb{R}^{n \times 1}$

Two rules of vector derivatives

There are two conventions in vector derivatives:

1. Gradient convention
2. Jacobian convention

Gradient convention

Jacobian convention

Derivative of a linear function

Derivative of a quadratic function

Back to Least square regression

In [46]:

```
n = salt_concentration_data.shape[0]
bfx = salt_concentration_data[:, 2:3]
bfy = salt_concentration_data[:, 1:2]
bfX = np.hstack((bfx, np.ones((bfx.shape[0], 1))))
bfX
```

Out[46]:

```
array([[0.19, 1.  ],
       [0.15, 1.  ],
       [0.57, 1.  ],
       [0.4 , 1.  ],
       [0.7 , 1.  ],
       [0.67, 1.  ],
       [0.63, 1.  ],
       [0.47, 1.  ],
       [0.75, 1.  ],
       [0.6 , 1.  ],
       [0.78, 1.  ],
       [0.81, 1.  ],
       [0.78, 1.  ],
       [0.69, 1.  ],
       [1.3 , 1.  ],
       [1.05, 1.  ],
       [1.52, 1.  ],
       [1.06, 1.  ],
       [1.74, 1.  ],
       [1.62, 1.  ]])
```

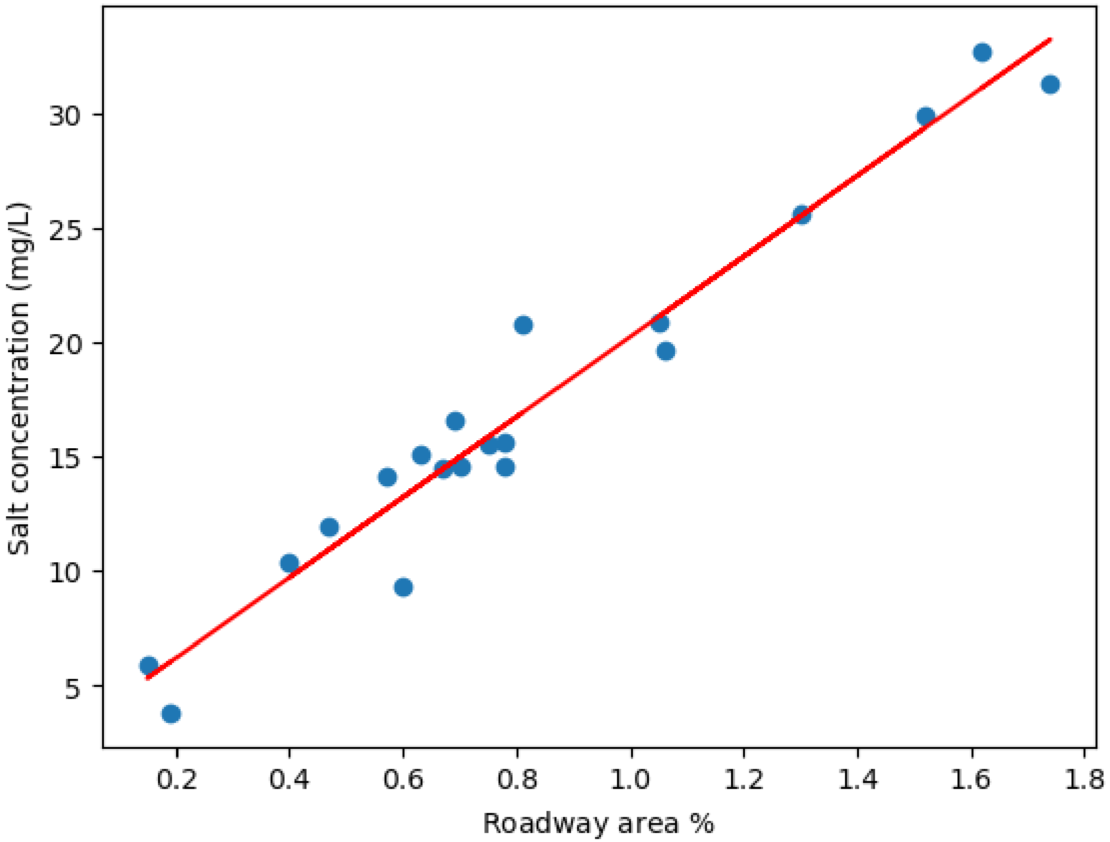
In [47]:

```
bfm = np.linalg.inv(bfX.T @ bfX) @ bfX.T @ bfy
print(bfm)
bfm, *_ = np.linalg.lstsq(bfX, bfy, rcond=None)
print(bfm)
```

```
[[17.5466671 ]
 [ 2.67654631]]
[[17.5466671 ]
 [ 2.67654631]]
```

In [48]:
```
m = bfm.flatten()[0]
c = bfm.flatten()[1]

# Plot the points
fig, ax = plt.subplots()
ax.scatter(salt_concentration_data[:, 2], salt_concentration_data[:, 1]
ax.set_xlabel(r"Roadway area $\%$")
ax.set_ylabel(r"Salt concentration (mg/L)")
x = salt_concentration_data[:, 2]
y = m * x + c
# Plot the points
ax.plot(x, y, 'r-') # the line
```

Out[48]:
```
[<matplotlib.lines.Line2D at 0x7fbf437f67c0>]
```


Salt concentration (mg/L)
Roadway area %
30
25
20
15
10
5
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8

Exercise 1

Derive the equations for least square linear regression when the equation of line is $\hat{\mathbf{w}}^\top \mathbf{x} + w_0 = 0$ instead of $y = mx + c$.

Hint: Convert the least square problem into equation of the form $\mathbf{v}^* = \arg\ \min_{\mathbf{v}} \|\mathbf{L}\mathbf{v}\|^2$ such that $\mathbf{v}^\top \mathbf{v} = 1$. Solve by finding null space of $\mathbf{L}$. $\mathbf{v}$ lies in the nullspace of $\mathbf{L}$. The nullspace of $\mathbf{L}$ is the last eigenvector (corresponding to the smallest eigenvalue) of $\mathbf{L}^\top \mathbf{L}$.

The error $e(x_i, y_i) = (y - (mx + c))^2$ can be visualized as distance of observed point from the fit line parallel to y-axis. Draw the visual for the errors of the form: $e(\mathbf{x}_i) = (\hat{\mathbf{w}}^\top \mathbf{x}_i + w_0 - 0)^2$. You do not need to use matplotlib. You can draw by hand or editing software.