Today

- Vectors, Matrices.
- Calculus
- Derivation with respect to a vector.
Linear Algebra

• Vectors
  – A one dimensional array.
  – If not specified, assume \( x \) is a column vector.

• Matrices
  – Higher dimensional array.
  – Typically denoted with capital letters.
  – \( n \) rows by \( m \) columns

\[
A = \begin{pmatrix}
  a_{0,0} & a_{0,1} & \cdots & a_{0,m-1} \\
  a_{1,0} & a_{1,1} & \cdots & a_{1,m-1} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n-1,0} & a_{n-1,1} & \cdots & a_{n-1,m-1}
\end{pmatrix}
\]

\[
x = \begin{pmatrix}
  x_0 \\
  x_1 \\
  \cdots \\
  x_{n-1}
\end{pmatrix}
\]
Transposition

- **Transposing** a matrix swaps columns and rows.

\[
x = \begin{pmatrix}
x_0 \\
x_1 \\
\vdots \\
x_{n-1}
\end{pmatrix}
\]

\[
x^T = \begin{pmatrix}
x_0 & x_1 & \ldots & x_{n-1}
\end{pmatrix}
\]
Transposition

- **Transposing** a matrix swaps columns and rows.

\[
A = \begin{pmatrix}
  a_{0,0} & a_{0,1} & \cdots & a_{0,m-1} \\
  a_{1,0} & a_{1,1} & \cdots & a_{1,m-1} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n-1,0} & a_{n-1,1} & \cdots & a_{n-1,m-1}
\end{pmatrix}
\]

\[
A^T = \begin{pmatrix}
  a_{0,0} & a_{1,0} & \cdots & a_{n-1,0} \\
  a_{0,1} & a_{1,1} & \cdots & a_{1,m-1} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{0,m-1} & a_{1,m-1} & \cdots & a_{n-1,m-1}
\end{pmatrix}
\]
Addition

• Matrices can be added to themselves iff they have the same dimensions.
  – A and B are both n-by-m matrices.

\[
A + B = \begin{pmatrix}
  a_{0,0} + b_{0,0} & a_{0,1} + b_{0,1} & \cdots & a_{0,m-1} + b_{0,m-1} \\
  a_{1,0} + b_{1,0} & a_{1,1} + b_{1,1} & & a_{1,m-1} + b_{1,m-1} \\
    \vdots & & & \vdots \\
  a_{n-1,0} + b_{n-1,0} & a_{n-1,1} + b_{n-1,1} & \cdots & a_{n-1,m-1} + b_{n-1,m-1}
\end{pmatrix}
\]
Multiplication

• To multiply two matrices, the inner dimensions must be the same.
  – An n-by-m matrix can be multiplied by an m-by-k matrix

\[ AB = C \]

\[ c_{ij} = \sum_{k=0}^{m} a_{ik} \times b_{kj} \]
Inversion

• The inverse of an n-by-n or square matrix A is denoted $A^{-1}$, and has the following property.

$$AA^{-1} = I$$

• Where $I$ is the identity matrix is an n-by-n matrix with ones along the diagonal.
  
  \[ I_{ij} = 1 \text{ iff } i = j, \ 0 \text{ otherwise} \]
Identity Matrix

- Matrices are invariant under multiplication by the identity matrix.

\[ AI = A \]
\[ IA = A \]
Helpful matrix inversion properties

\[(A^{-1})^{-1} = A\]

\[(kA)^{-1} = k^{-1} A^{-1}\]

\[(A^T)^{-1} = (A^{-1})^T\]

\[(AB)^{-1} = B^{-1} A^{-1}\]
Norm

- The norm of a vector, $x$, represents the euclidean length of a vector.

\[
\|x\| = \sqrt{\sum_{i=0}^{n-1} x_i^2}
\]

\[
= \sqrt{x_0^2 + x_1^2 + \ldots + x_{n-1}^2}
\]
Positive Definite-ness

• Quadratic form
  – Scalar \[ c_0 + c_1 x + c_2 x^2 \]
  – Vector \[ x^T A x \]

• Positive Definite matrix \( M \)
  \[ x^T M x > 0 \]

• Positive Semi-definite
  \[ x^T M x \geq 0 \]
Calculus

- Derivatives and Integrals
- Optimization
Derivatives

• A **derivative** of a function defines the slope at a point $x$. 
Derivative Example
Integrals

- **Integration** is the inverse operation of derivation (plus a constant)

\[ \int f(x) \, dx = F(x) + c \]

- Graphically, an integral can be considered the area under the curve defined by \( f(x) \)
Integration Example

\[ \int_{a}^{b} f(x) \, dx \]

\[ \sum f(x) \, dx \]
Vector Calculus

• Derivation with respect to a matrix or vector
• Gradient
• Change of Variables with a Vector
Derivative w.r.t. a vector

• Given a vector $\mathbf{x}$, and a function $f(\mathbf{x})$, how can we find $f'(\mathbf{x})$?

\[ f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R} \]
Derivative w.r.t. a vector

- Given a vector $\mathbf{x}$, and a function $f(\mathbf{x})$, how can we find $f'(\mathbf{x})$?

\[
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix}
\frac{\partial f(\mathbf{x})}{\partial x_0} \\
\frac{\partial f(\mathbf{x})}{\partial x_1} \\
\vdots \\
\frac{\partial f(\mathbf{x})}{\partial x_{n-1}}
\end{pmatrix}
\]

$f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$
Example Derivation

\[ f(\vec{x}) = x_0 + 4x_1 x_2 \]

\[
\frac{\partial f(\vec{x})}{\partial x_0} = 1
\]

\[
\frac{\partial f(\vec{x})}{\partial x_1} = 4x_2
\]

\[
\frac{\partial f(\vec{x})}{\partial x_2} = 4x_1
\]
Example Derivation

\[ f(\vec{x}) = x_0 + 4x_1 x_2 \]

\[
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix}
\frac{\partial f(\mathbf{x})}{\partial x_0} \\
\frac{\partial f(\mathbf{x})}{\partial x_1} \\
\frac{\partial f(\mathbf{x})}{\partial x_2}
\end{pmatrix} = \begin{pmatrix}
1 \\
4x_2 \\
4x_1
\end{pmatrix}
\]

Also referred to as the gradient of a function.

\[ \nabla f(\mathbf{x}) \text{ or } \nabla f \]
Matrix derivative

- Given two vectors $y$ and $x$, how can we find $y'$?

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad \quad \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$
Useful Vector Calculus identities

• Scalar Multiplication
  \[
  \frac{\partial}{\partial \vec{x}} (\vec{x}^T \vec{a}) = \frac{\partial}{\partial \vec{x}} (\vec{a}^T \vec{x}) = \vec{a}
  \]

• Product Rule
  \[
  \frac{\partial}{\partial \vec{x}} (AB) = \frac{\partial A}{\partial \vec{x}} B + A \frac{\partial B}{\partial \vec{x}}
  \]
  \[
  \frac{\partial}{\partial \vec{x}} (\vec{x}^T A) = A
  \]
  \[
  \frac{\partial}{\partial \vec{x}} (A\vec{x}) = A^T
  \]
Useful Vector Calculus identities

- Derivative of an inverse
  \[ \frac{\partial}{\partial x} (A^{-1}) = -A^{-1} \frac{\partial A}{\partial x} A^{-1} \]

- Change of Variable
  \[ \int f(\vec{x}) d\vec{x} = \int f(\vec{u}) \left| \frac{\partial \vec{x}}{\partial \vec{u}} \right| d\vec{u} \]
Optimization

• Have an objective function that we’d like to maximize or minimize, \( f(x) \)

• Set the first derivative to zero.
Optimization with constraints

• What if I want to constrain the parameters of the model.
  – The mean is less than 10

• Find the best likelihood, subject to a constraint.

• Two functions:
  – An objective function to maximize
  – An inequality that must be satisfied
Lagrange Multipliers

- Find maxima of $f(x,y)$ subject to a constraint.

$$f(x, y) = x + 2y$$

$$x^2 + y^2 = 1$$
General form

- Maximizing: \( f(x, y) \)
- Subject to: \( g(x, y) = c \)

- Introduce a new variable, and find a maxima.

\[
\Lambda(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c)
\]
Example

• Maximizing: $f(x, y) = x + 2y$
• Subject to: $x^2 + y^2 = 1$
Why does Machine Learning need these tools?

• Calculus
  – We need to identify the maximum likelihood, or minimum risk. Optimization
  – Integration allows the marginalization of continuous probability density functions

• Linear Algebra
  – Many features leads to high dimensional spaces
  – Vectors and matrices allow us to compactly describe and manipulate high dimensional feature spaces.
Why does Machine Learning need these tools?

• Vector Calculus
  – All of the optimization needs to be performed in high dimensional spaces
  – Optimization of multiple variables simultaneously – Gradient Descent
  – Want to take a marginal over high dimensional distributions like Gaussians.
Next Time

• Linear Regression
  – Then Regularization