

LaTeX Equation Practice

Raymond Heberer
ray.heberer@gmail.com

Friday 25th May, 2018

1 Introduction

Ideas from mathematics underly virtually every technique and concept in Machine Learning (ML). While possessing a rigorous understanding of all these ideas is certainly not required, it can be beneficial. For the beginning practitioner, the scope of the math that forms the foundation of ML is intimidating, and many choose to reassure themselves by observing that modern programming frameworks mean that the ability to implement an algorithm is far removed from the ability to understand its mathematical underpinnings.

This "top-down" approach, emphasizing results over theory, is both effective and marketable. However, there are certain mathematical foundations of ML that are worth becoming familiar with.

In what follows, I state some of the most important equations a beginner should seek to understand when approaching ML. The contents of this document will lay the groundwork for at least one article on Medium.

2 Functions and Derivatives

2.1 Variables and Function Notation

$$y = f(x)$$

2.2 Equation of a Line

$$y = mx + b$$

2.3 The Power Rule

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

2.4 Finite Differences and Difference Quotients

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

$$\begin{aligned}\Delta f(x) &= f(x + \Delta x) - f(x) \\ \nabla f(x) &= f(x) - f(x - \Delta x) \\ \delta f(x) &= f\left(x + \frac{1}{2}\Delta x\right) - f\left(x - \frac{1}{2}\Delta x\right)\end{aligned}$$

3 Vectors and Matrices

$$\vec{a}, \vec{b} \in \mathbb{R}^n$$

3.1 Vector Arithmetic

$$\begin{aligned}\vec{a} + \vec{b} &= (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) \\ \lambda \vec{a} &= (\lambda a_1, \lambda a_2, \dots, \lambda a_n)\end{aligned}$$

3.2 Vector Magnitude

$$|\vec{a}| = \sum_{i=1}^n \sqrt{a_i^2} = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

3.3 L1 and L2 Norms

$$\begin{aligned}||\vec{a}||_1 &= \sum_{i=1}^n |a_i| \\ ||\vec{a}||_2 &= \sum_{i=1}^n \sqrt{a_i^2}\end{aligned}$$

3.4 Matrix Multiplication

$$\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$\mathbf{AB} \in \mathbb{R}^{m \times p}$$

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mp} \end{pmatrix}$$

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{im}b_{mj} = \sum_{k=1}^m a_{ik}b_{kj}$$

3.5 Matrix Inversion

$$\mathbf{A} \in \mathbb{R}^{n \times n}$$

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$$

$$\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$$

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

4 Statistical Relationships and Procedures

4.1 Correlation

$$\rho_{\mathbf{X}, \mathbf{Y}} = \frac{\text{cov}(\mathbf{X}, \mathbf{Y})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}$$

$$= \frac{E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})]}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

4.2 Principal Components Analysis

$$\mathbf{w}_{(1)} = \arg \max_{\|\mathbf{w}\|=1} \{ \|\mathbf{X}\mathbf{w}\|^2 \} = \arg \max_{\|\mathbf{w}\|=1} \{ \|\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}\|^2 \}$$

$$\mathbf{w}_{(1)} = \arg \max \left\{ \frac{\|\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}\|^2}{\mathbf{w}^T \mathbf{w}} \right\}$$

$$\mathbf{W}_k = \arg \max_{\|\mathbf{w}\|=1} \left\{ \|\hat{\mathbf{X}}_k \mathbf{w}\|^2 \right\}$$

$$\hat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X} \mathbf{w}_{(s)} \mathbf{w}_{(s)}^T$$

4.2.1 Singular-Value Decomposition (SVD)

$$\mathbf{M} \in \mathbb{R}^{m \times n}$$

$$\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$$

- The nonzero singular values are the square roots of the nonzero eigenvalues of $\mathbf{M}^* \mathbf{M}$ or $\mathbf{M} \mathbf{M}^*$
- $\mathbf{\Sigma}$ is a diagonal $(m \times n)$ matrix of non-negative real numbers, known as the **singular values** of \mathbf{M}
- The columns of \mathbf{V} (right singular vectors) are eigenvectors of $\mathbf{M}^* \mathbf{M}$
- The columns of \mathbf{U} (left singular vectors) are eigenvectors of $\mathbf{M} \mathbf{M}^*$