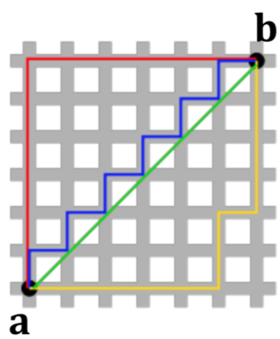
Contents

1 Regression

1.1 Warmup

- An n x d matrix is one that has n rows and d columns
- Vector Norms:
  - l_p norm: $||x||_p = (\sum_i |x_i|^p)^{1/p}$
  - l_2 norm (Euclidean Norm): $||x||_2 = (\sum_i x_i^2)^{1/2}$
  - l_∞ norm: $||x||_\infty = \max_i |x_i|$



- The ℓ_2-distance (Euclidean distance): $||\mathbf{a} - \mathbf{b}||_2$ (green line)
- The ℓ_1-distance (Manhattan distance): $||\mathbf{a} - \mathbf{b}||_1$ (red, blue, yellow lines)

Figure 1: Norm distances

1.2 Rank

- Rank: The number of linearly independent rows (or columns).
- Full Rank: a square matrix is full rank if the rank equals to #columns.

1.3 Eigenvalue Decomposition

- let be any $\times$ symmetric matrix.
- Eigenvalue decomposition: $\mathrm{A} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_{\mathrm{i}} \mathrm{v}_{\mathrm{i}} \mathrm{v}_{\mathrm{i}}^{\mathrm{T}}$
- Eigenvalues satisfy $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{\mathrm{n}}|$
- Eigenvectors satisfy $\mathrm{v}_{\mathrm{i}}^{\mathrm{T}} \mathrm{v}_{\mathrm{j}} = 0$ for all $\mathrm{i} \neq \mathrm{j}$
- **A** is a full rank iff all the eigenvalues are nonzero

1.4 Least Squares Regression with Gradient Descent

1. Least squares regression model: x{}