Probability and Linear Algebra Review

Jin Sun
Henry Gifford
Probability

• Random Variables
• Independence
• Expectation and Variance
• Distributions
Random Variables

• Sample space $\Omega$
• Random variable is a function mapping $X: \Omega \rightarrow \mathbb{R}$
• Probability: $P(X = x) = \sum_{\omega \in \Omega \mid X(\omega) = x} P_\omega$
  • Marginal probability
  • Conditional probability
  • Joint probability
• $P(X)$ is a function. $P(X = x)$ is a real value.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>$P(X, Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Independence

• Marginal Independence
  • $P(X, Y) = P(X)P(Y)$
  • $P(X|Y) = P(X)$

• Conditional Independence
  • $P(X, Y | C) = P(X | C)P(Y | C)$
  • $P(X | Y, C) = P(X | C)$

• Marginal Independence does not infer conditional independence, vice versa. (d-separation, later in this course)

• Chain rule:
**Expectation**

• **Definition**
  - \( E(X) = \sum_{x \in X} xP(X = x) \)
  - \( E(X) = \int_{-\infty}^{\infty} xf_X(x)dx \)

• **Conditional Expectation**
  - \( E(X | Y = y) = \sum_{x \in X} xP(X = x | Y = y) = \frac{\sum_{x \in X} xP(X=x,Y=y)}{P(Y=y)} \)
  - \( E(X | Y = y) = \int_{X} xf_X(x | Y = y) dx \)
Variance and Covariance

• Variance
  • $Var(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$
  • Equal likely case: $Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$
    • Not $\frac{1}{n-1}$. Be careful with Matlab.

• Covariance
  • $\sigma(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$
Distribution

• Multivariate Gaussian Distribution

\[ f_X(x_1, x_2, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right) \]

• \( \Sigma \) – Covariance Matrix \( \mathcal{R}^{k \times k} \), symmetric positive semi definite

• How many parameters required to estimate this distribution?
  • Mean – \( k \)
  • Covariance – \((1+k)*k/2\)
Linear Algebra

• Property of Matrices
• Transpose and Inverse
• Vector Norms
• Matrix Calculus
Property of Matrices

• Associative: \((AB)C = A(BC)\)
• Distributive: \(A(B + C) = AB + AC\)
• **NOT** commutative: \(AB \neq BA\)

• In matrix multiplication, inner dimensions must agree.

\[
A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times q}, A \times B \in \mathbb{R}^{m \times q}, B \times A \text{ does not exist}
\]
Transposer and Inverse

\[(R_{m \times n})^T = R_{n \times m}\]  
\[(R_{m \times m})^{-1} = R_{m \times m}\]

For square matrix:

- \((A^T)^T = A\)
- \((AB)^T = B^T A^T\)
- \((A + B)^T = A^T + B^T\)
- \((A^{-1})^{-1} = A\)
- \((AB)^{-1} = B^{-1} A^{-1}\)
- \((A^T)^{-1} = (A^{-1})^T\)

Pseudo inverse for rectangular matrices \((R_{m \times n})^{-1} = R_{n \times m}\)
Vector Norms

• Family of norms
  \[ ||x||_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \]

• \( l_1 \) norm: \( ||x||_1 = \sum_{i=1}^{n} |x_i| \)

• \( l_2 \) norm: \( ||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \)
Matrix Calculus

• Some useful sources
  http://www.cs.cmu.edu/~10601b/resource/linearalgebra.pdf
  http://www.cs.nyu.edu/~roweis/notes/matrixid.pdf
Derive normal equation by yourself

• Solve \( \arg\min_w L(w) = \arg\min_w \frac{1}{2} ||y - Xw||_2^2 \)

Where \( y \in \mathcal{R}^{n*1}, X \in \mathcal{R}^{n*f}, w \in \mathcal{R}^{f*1} \)

\[
\frac{\partial (a^T x)}{\partial x} = \frac{\partial (x^T a)}{\partial x} = a \\
\frac{\partial (x^T Ax)}{\partial x} = (A + A^T)x \\
\frac{\partial (a^T X b)}{\partial X} = ab^T \\
\frac{\partial (a^T X^T b)}{\partial X} = ba^T \\
\frac{\partial (a^T X a)}{\partial X} = \frac{\partial (a^T X^T a)}{\partial X} = aa^T \\
\frac{\partial (x^T A)}{\partial x} = A \\
\frac{\partial (x^T)}{\partial x} = I \\
\frac{\partial (A x)}{\partial z} = A \frac{\partial x}{\partial z} \\
\frac{\partial (X Y)}{\partial z} = X \frac{\partial Y}{\partial z} + \frac{\partial X}{\partial z} Y \\
\frac{\partial (X^{-1})}{\partial z} = -X^{-1} \frac{\partial X}{\partial z} X^{-1} \\
\frac{\partial \ln |X|}{\partial X} = (X^{-1})^T = (X^T)^{-1}
\]
Solution

• \( L(w) = \frac{1}{2} ||y - Xw||^2_2 = \frac{1}{2} (y - Xw)^T (y - Xw) = \frac{1}{2} (y^T y - y^T Xw - w^T X^T y + w^T X^T X w) \)

• \( \frac{\partial L}{\partial w} = \frac{1}{2} (0 - X^T y - X^T y + 2X^T X w) = -X^T y + X^T X w \)

• Set derivative to zero: \( X^T X w = X^T y \)

• \( w = (X^T X)^{-1} X^T y \)