1 Identities with Expectation

1.1

When $n = k$:

$$E[X^k] = \frac{k!}{\lambda^k} = \int_0^\infty x^k f(x)dx = \lambda \int_0^\infty x^k e^{-\lambda x} dx$$

When $n = k + 1$:

$$\begin{aligned} E[X^{k+1}] &= \int_0^\infty x^{k+1} f(x)dx \\ &= \lambda \int_0^\infty x^{k+1} e^{-\lambda x} dx \\ &= -\int_0^\infty x^{k+1} de^{-\lambda x} \\ &= -[x^{k+1} e^{-\lambda x}|_0^\infty - \int_0^\infty (k+1)x^k e^{-\lambda x} dx] = \frac{(k+1)!}{\lambda^{k+1}} \end{aligned}$$

1.2

$$\begin{aligned} E[X] &= \int_0^\infty x f(x)dx \\ &= \int_0^\infty x dF(x) \\ &= -\int_0^\infty x d(1 - F(x)) \\ &= -x(1 - F(x))|_0^\infty + \int_0^\infty (1 - F(x))dx \\ &= \int_0^\infty P(X \geq t)dt \end{aligned}$$

1.3

$$\begin{aligned} E[X] &= E[X\mathbf{1}\{X = 0\} + E[X\mathbf{1}\{X > 0\} \\ &= 0 + E[X\mathbf{1}\{X > 0\} \\ &\leq \sqrt{E[X^2]E[(\mathbf{1}\{X > 0\})^2]} && (Cauchy\text{–}Schwarz) \\ &= \sqrt{E[X^2]P(X > 0)} && (P(A) = E[\mathbf{1}\{A\}]) \end{aligned}$$

1.4

$$\begin{aligned} E[t - X] &\leq E[(t - X)\mathbf{1}\{t - X > 0\}] \\ &\leq \sqrt{E[t^2 + X^2]E[(\mathbf{1}\{t - X > 0\})^2]} \\ &= \sqrt{(t^2 + E[X^2])(1 - P(X \geq t))} \end{aligned}$$

2 Probability Potpourri

2.1

$$\begin{aligned} x^T \Sigma x &= x^T E[(Z-\mu)(Z-\mu)^T]x \\ &= E[x^T(Z-\mu)(Z-\mu)^T x] \\ &= E[[(Z-\mu)^T x]^T [(Z-\mu)^T x]] \\ &= E[||(Z-\mu)^T x||^2] \\ &\geq 0 \end{aligned}$$

2.2

Define:

$$H : hit$$

$$W : windy$$

Given:

$$P(H|W) = 0.4$$

$$P(H|\overline{W}) = 0.7$$

$$P(W) = 0.3$$

(i)

$$P(H|W) = 0.4$$

(ii)

$$P(H) = P(HW) + P(H\overline{W}) = p(W)P(H|W) + P(\overline{W})P(H|\overline{W})$$

(iii)

$$2P(H)(1-P(H))$$

(iv)

$$P(\overline{W}|\overline{H}) = \frac{P(\overline{W})P(\overline{H}|\overline{W})}{p(W)P(\overline{H}|W) + P(\overline{W})P(\overline{H}|\overline{W})}$$

2.3

$$E[score] = \int_0^{\frac{1}{\sqrt{3}}} 4f(x)dx + \int_{\frac{1}{\sqrt{3}}}^1 3f(x)dx + \int_1^{\sqrt{3}} 2f(x)dx$$

2.4

$$
\begin{aligned}
P(X=k|X+Y=n) &= \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)} \\
&= \frac{\frac{\lambda^k}{k!}e^{-\lambda}\frac{\mu^{n-k}}{(n-k)!}e^{-\lambda}}{\frac{(\lambda+\mu)^n}{n!}e^{-\lambda}} \\
&= C_n^k\frac{\lambda^k\mu^{n-k}}{(\lambda+\mu)^k}e^{-\lambda}
\end{aligned}
$$

3 Properties of Gaussians

3.1

$$
\begin{aligned}
E[e^{-\lambda X}] &= \int_{-\infty}^{+\infty} e^{-\lambda X}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} e^{-(\frac{x}{\sqrt{2}\sigma}-\frac{\sigma\lambda}{\sqrt{2}})^2+\frac{\sigma^2\lambda^2}{2}}\,dx \\
&= \frac{e^{\frac{\sigma^2\lambda^2}{2}}}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-t^2}\,dt \\
&= e^{\frac{\sigma^2\lambda^2}{2}}
\end{aligned}
$$

3.2

$$
P(X\geq t)=P(e^{\lambda X}\geq e^{\lambda t})\leq\frac{E[e^{\lambda X}]}{e^{\lambda t}},\lambda=\frac{t}{\sigma^2}
$$

3.3

$$
\sum_{i=1}^{n}X_i\sim N(0,n\sigma^2)
$$

$$
\frac{1}{n}\sum_{i=1}^{n}X_i\sim N(0,\frac{\sigma^2}{n})
$$

3.4

$$
X\sim N(0,1),Y\sim\begin{cases} X, & p=0.5 \\ -X, & 1-p=0.5\end{cases}
$$

3.5

$$
u_x\sim N(0,\sum u_i^2)
$$

$$
v_x\sim N(0,\sum v_i^2)
$$

$$
\sum u_iv_i=0
$$

$$
Cov(u_x,v_x)=\frac{1}{2}[D(X+Y)-DX-DY]=\sum(u_i+v_i)^2-\sum u_i^2-\sum v_i^2=0
$$

3.6

???

4 Linear Algebra Review

4.1

(a)

$$x_T Ax \geq 0$$

(b)

$$Ax_i = \lambda_i x_i$$

$$x_i^T Ax_i = \lambda_i ||x_i||^2 \geq 0$$

$$\lambda_i \geq 0$$

(c)

$$A = Q\Lambda Q^T, \Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_m, 0, \cdots, 0)$$

$$B = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_m}, 0, \cdots, 0)$$

$$U = QB^T, U^T = BQ^T$$

$$A = UU^T$$

4.2

(a)

$$x^T(2A + 3B)x = 2x^T Ax + 3x^T Bx \geq 0$$

(b)

$$e_k = (0, 0, \cdots, 1, 0, \cdots, 0)^T, A_{kk} = e_k^T Ae_k \geq 0$$

(c)

$$e = (1, 1, \cdots, 1)^T, \sum_i \sum_j A_{ij} = e^T Ae \geq 0$$

(d) ???

(e) ???

4.3

eigendecomposition

4.4

Step 1: All Real Symmetric Matrices can be diagonalized in the form: $H = Q\Lambda Q^T$ So, $v^T Hv = v^T Q\Lambda Q^T v$.

Step 2: Define transformed vector: $y = Q^T v$. So, $v^T Hv = y^T \Lambda y$.

Step 3: Expand $y^T \Lambda y = \lambda_{max} y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_{min} y_N^2 \leq \lambda_{max} y^T y = \lambda_{max} v^T v$.

Step 6: Putting it all back together. $v^T Hv \leq \lambda_{max} v^T v = \lambda_{max}$.

5 Gradients and Norms

5.1

$$||x||_\infty \leq ||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2 \leq n||x||_\infty$$

Cauchy-Schwarz inequality:

$$y = (1, 1, \cdots, 1)^T, | < x, y > | = ||x||_1 \leq ||x||_2 ||y||_2 = \sqrt{n}||x||_2$$

5.2

(a)

$$\frac{y_i}{\beta_i}$$

(b)

$$0$$

(c)

$$A_{ij}$$

(d)

$$\sum_{i=1}^{n} \frac{y_i Aij}{\tanh(Ax + b)_1}$$

5.3

pass

5.4

pass

5.5

$$\begin{aligned} L(\theta) &= ||y - X\theta||_2^2 \\ &= (y - X\theta)^T (y - X\theta) \\ &= y^T y + \theta^T X^T X\theta - 2y^T X\theta \end{aligned}$$

$$\begin{aligned} \nabla L(\theta) &= 2X^T X\theta - 2X^T y \\ &= 0 \end{aligned}$$

$$\theta^* = (X^T X)^{-1} X^T y$$

6 Gradient Descent

6.1

$$L(x) = \frac{1}{2}x^T Ax - b^T x$$

$$\nabla L(x) = \frac{1}{2}(A + A^T)x - b = 0$$

$$x^* = 2(A + A^T)^{-1}b = A^{-1}b$$

6.2

let $i = 0$, $x^{(0)}$<=some arbitrary vector

`while True`:

$x^{(i+1)} = x^{(i)} - \nabla f(x^{(i)})$.

`if` $||x^{(i+1)} - x^{(i)}|| < \epsilon$:

`break`.

$i = i + 1$

$x^* = x^{(i+1)}$

6.3

Using 6.1 & 6.2

6.4

$$||Ax||_2 \leq ||A||_2||x||_2 = \sqrt{\lambda_{max(A^T A)}}||x||_2 = \sqrt{\lambda_{max(A^2)}}||x||_2 = \lambda_{max(A)}||x||_2$$

6.5

$||x^{(k)} - x^*||_2 = ||I - A||_2||x^{(k-1)} - x^*||_2 \leq \lambda_{max(I-A)}||x^{(k-1)} - x^*||_2 = (1 - \lambda_{max(A)})||x^{(k-1)} - x^*||_2$

6.6

$$k \leq \log_\rho \frac{\epsilon}{||x^{(0)} - x^*||_2}$$