Homework 2

CS 189 Spring 2019 Introduction to Machine Learning

1 Identities with Expectation

Q1

> If $u = u(x)$ and $du = u'(x)dx$, while $v = v(x)$ and $dv = v'(x)dx$, then the integration by parts formula states that
>
> $$\int udv = uv - \int vdu$$

Base case: $E[X^0] = 1$. Inductive hypothesis: $E[X^k] = \frac{k!}{\lambda^k}$. Inductive step:

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

QED.

Q2

Since X is a non-negative real-valued random variable, we have

$$
\begin{aligned}
E[X] &= E[\int_0^\infty \mathbf{1}\{X \geq t\}dt] \\
&= \int_0^\infty \int_0^\infty \mathbf{1}\{x \geq t\}dt f(x)dx \\
&= \int_0^\infty \int_0^\infty \mathbf{1}\{x \geq t\}f(x)dxdt
\end{aligned}
$$

Note that $\Pr(X \geq t) = E[\mathbf{1}\{X \geq t\}]$, so $E[X] = \int_0^\infty \Pr(X \geq t)dt$. **QED.**

Q3

Since $X \geq 0$, we have $X = X\mathbf{1}\{X > 0\}$. Now, we apply Cauchy-Schwarz inequality:

$$
\begin{aligned}
(E[X])^2 &= (E[X\mathbf{1}\{X > 0\}])^2 \\
&\leq E[X^2]E[(\mathbf{1}\{X > 0\})^2] \\
&= E[X^2]E[\mathbf{1}\{X > 0\}] \\
&= E[X^2]\Pr(X > 0)
\end{aligned}
$$

In conclusion, $\Pr(X > 0) \geq \frac{(E[X])^2}{E[X^2]}$. **QED.**

Q4

Use the fact that $t - X \leq (t - X)\mathbf{1}\{t - X > 0\}$, and apply Cauchy-Schwarz inequality:

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

Note that $E[X] = 0$, so the LHS is t^2, and the first expectation in the RHS is $t^2 + E[X^2]$. The second one is $\Pr(t > X)$.

Therefore, we have $t^2 \leq (t^2 + E[X^2])\Pr(t > X)$, which after some rearranging gives $\Pr(X \geq t) \leq \frac{E[X^2]}{E[X^2]+t^2}$, as desired. **QED.**

2 Properties of Gaussians

Q1

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

QED.

Q2

> The **Chernoff bound** for a random variable X is obtained by applying Markov's inequality to $e^{\lambda X}$.

> **Markov's inequality** If X is a nonnegative random variable and $a > 0$, then the probability that X is at least a is at most the expectation of X divided by a. $\Pr(X \geq a) \leq \frac{E[X]}{a}$.

For any $\lambda > 0$ and $t > 0$, $\Pr(X \geq t) = \Pr(e^{\lambda X} \geq e^{\lambda t}) \leq \frac{E[e^{\lambda X}]}{e^{\lambda t}} = e^{\sigma^2\lambda^2/2 - \lambda t}$. Setting $\lambda = \frac{t}{\sigma^2}$ gives the result. Similar for $\lambda < 0$. **QED.**

Q3

Let $X = \frac{1}{n}\sum_{i=1}^n X_i$. Since X_is are independent randoms variables, we have $E[X] = \frac{1}{n}\sum_{i=1}^n E[X_i] = 0$ and $\mathrm{Var}[X] = \sum_{i=1}^n \mathrm{Var}[\frac{X_i}{n}] = n \cdot \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$. Therefore, $X \sim N(0, \frac{\sigma^2}{n})$. $\Pr(X \geq t) \leq e^{-\frac{t^2 n}{2\sigma^2}}$. If $n \to \infty$, then the probabilty of X being away from 0 goes to 0.

Q4

Let $X \sim N(0,1)$ and ξ be a random variable such that $\Pr(\xi = -1) = \Pr(\xi = 1) = 0.5$. Let $Y = \xi \cdot X$. Note that $Y \sim N(0,1)$. However, $X + Y$ is not Gaussian.

Q5

$$E[u_x v_x] = E[(\sum_{i=1}^{n} u_i X_i)(\sum_{i=1}^{n} v_i X_i)]$$

Note that if $\mathbf{Cov}(X_i X_j) = 0$ if $i \neq j$, which implies $E[X_i X_j] = E[X_i]E[X_j] = 0$. Therefore, $E[u_x v_x] = \sum_{i=1}^{n} u_i v_i E[X_i^2] = \langle u_i, v_i \rangle = 0$. Since Gaussian random variables are independent iff they are uncorrelated, we conclude that u_x and v_x are independent.

Q6

$$\begin{aligned}
\lambda E\left[\max_i |X_i|\right] &= \log \exp\left\{E\left[\lambda \max_i |X_i|\right]\right\} \\
&\leq \log E\left[\exp\left\{\lambda \max_i |X_i|\right\}\right] \\
&\leq \log E\left[\sum_i \exp\left\{\lambda |X_i|\right\}\right] \\
&\leq \log E\left[\sum_i \left(\exp\left\{\lambda X_i\right\} + \exp\left\{-\lambda X_i\right\}\right)\right] \\
&= \log \sum_i \left(E[e^{\lambda X_i}] + E[e^{-\lambda X_i}]\right) \\
&= \log \sum_i (2e^{\sigma^2 \lambda^2 / 2}) \\
&= \log 2n + \sigma^2 \lambda^2 / 2
\end{aligned}$$

Setting $\lambda = \frac{\sqrt{\log 2n}}{\sigma}$ gives $E\left[\max_i |X_i|\right] \leq \frac{3}{2}\sqrt{\log 2n}\sigma$. **QED.**

3 Linear Algebra Review

Q1

$(a) \Rightarrow (b)$ For all the eigenvectors $v_i's$ of A with the corresponding eigenvalues $\lambda_i's$. We have $v_i^\top A v_i = \lambda_i v_i^\top v_i = \lambda_i ||v_i||^2 \geq 0$. Therefore, all the eigenvalues of A are non-negative.

Q2

Q3

4 Gradients and Norms

5 Covariance Practice