Statistics for high and ∞–dimensional data

Exercise 1.1

Lasse Vuursteen

February 13, 2018

Exercise 1.1

Let f, g be positive probability densities with respect to some measure μ. Show that $\mathrm{KL}(f, g) \geq 0$ and that $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Exercise 1.1

Let f, g be positive probability densities with respect to some measure μ. Show that $\mathrm{KL}(f, g) \geq 0$ and that $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:

Exercise 1.1

Let f, g be positive probability densities with respect to some measure μ. Show that $\mathrm{KL}(f, g) \geq 0$ and that $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:

Let $\mathrm{KL}(f, g)$ be defined as on page 5. That is,

$$\mathrm{KL}(f, g) := \int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu$$

Exercise 1.1

Let f, g be positive probability densities with respect to some measure μ. Show that $\mathrm{KL}(f, g) \geq 0$ and that $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:

Let $\mathrm{KL}(f, g)$ be defined as on page 5. That is,

$$\mathrm{KL}(f, g) := \int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu$$

This is well defined if $\mu(\{g = 0\}) = 0$.

Showing $\mathrm{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$.
Define the distribution $F(E) := \int_E f \, \mathrm{d}\mu, \; E \in \mathcal{F}$.

Showing $\mathrm{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$.
Define the distribution $F(E) := \int_E f \, \mathrm{d}\mu, \ E \in \mathcal{F}$.
Write

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int \log\left(\frac{f}{g}\right) \mathrm{d}F = \int -\log\left(\frac{g}{f}\right) \mathrm{d}F$$

Showing $\mathrm{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$.
Define the distribution $F(E) := \int_E f \,\mathrm{d}\mu, \; E \in \mathcal{F}$.
Write

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int \log\left(\frac{f}{g}\right) \mathrm{d}F = \int -\log\left(\frac{g}{f}\right) \mathrm{d}F$$

Note that the mapping $\mathbb{R} \to \mathbb{R}, x \mapsto -\log(x)$ is convex.

Showing $\mathrm{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$.
Define the distribution $F(E) := \int_E f \, \mathrm{d}\mu, \;\; E \in \mathcal{F}$.
Write

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int \log\left(\frac{f}{g}\right) \mathrm{d}F = \int -\log\left(\frac{g}{f}\right) \mathrm{d}F$$

Note that the mapping $\mathbb{R} \to \mathbb{R}, x \mapsto -\log(x)$ is convex.
By Jensen's inequality,

$$\int -\log\left(\frac{g}{f}\right) \mathrm{d}F \geq -\log \int \left(\frac{g}{f}\right) \mathrm{d}F = -\log \int f \left(\frac{g}{f}\right) \mathrm{d}\mu = 0.$$

Showing $\mathrm{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$.
Define the distribution $F(E) := \int_E f \, \mathrm{d}\mu, \; E \in \mathcal{F}$.
Write

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int \log\left(\frac{f}{g}\right) \mathrm{d}F = \int -\log\left(\frac{g}{f}\right) \mathrm{d}F$$

Note that the mapping $\mathbb{R} \to \mathbb{R}, x \mapsto -\log(x)$ is convex.
By Jensen's inequality,

$$\int -\log\left(\frac{g}{f}\right) \mathrm{d}F \geq -\log \int \left(\frac{g}{f}\right) \mathrm{d}F = -\log \int f \left(\frac{g}{f}\right) \mathrm{d}\mu = 0.$$

Conclude: $\mathrm{KL}(f, g) \geq 0$.

Showing $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–a.e.

Easy part: " $\Longleftarrow$:" Let $f = g$ μ–a.e.. Then

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int f \log\left(\frac{f}{g}\right) \mathbb{1}_{\{f=g\}} \mathrm{d}\mu = 0.$$

More difficult: " $\Longrightarrow$ " part. Following the hint, it will be shown that

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu \geq \int (\sqrt{f} - \sqrt{g})^2 \mathrm{d}\mu = d_{H^2}(f, g)$$

Showing $\mathrm{KL}(f, g) = 0 \iff f = g$ μ–a.e.

Easy part: " $\Longleftarrow$:" Let $f = g$ μ–a.e.. Then

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int f \log\left(\frac{f}{g}\right) \mathbb{1}_{\{f=g\}} \mathrm{d}\mu = 0.$$

More difficult: " $\Longrightarrow$ " part. Following the hint, it will be shown that

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu \geq \int (\sqrt{f} - \sqrt{g})^2 \mathrm{d}\mu = d_{H^2}(f, g)$$

$d_{H^2}(f, g)$ is the metric called squared Hellinger distance.

Showing $\mathrm{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int -f \log\left(\frac{g}{f}\right) \mathrm{d}\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \mathrm{d}\mu$$
$$= 1 + 1 - 2 \int \sqrt{f}\sqrt{g}\, \mathrm{d}\mu$$

Showing $\mathrm{KL}(f,g) \geq d_{H^2}(f,g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\begin{aligned}
\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu &= \int -f \log\left(\frac{g}{f}\right) \mathrm{d}\mu \geq \int f\left(2 - 2\sqrt{\frac{g}{f}}\right) \mathrm{d}\mu \\
&= 1 + 1 - 2\int \sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\
&= \int f + g - 2\sqrt{f}\sqrt{g}\, \mathrm{d}\mu
\end{aligned}$$

Showing $\mathrm{KL}(f,g) \geq d_{H^2}(f,g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\begin{aligned}\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu &= \int -f \log\left(\frac{g}{f}\right) \mathrm{d}\mu \geq \int f\left(2 - 2\sqrt{\frac{g}{f}}\right) \mathrm{d}\mu \\ &= 1 + 1 - 2\int \sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\ &= \int f + g - 2\sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\ &= \int (\sqrt{f} - \sqrt{g})^2\, \mathrm{d}\mu = d_{H^2}(f,g)\end{aligned}$$

Showing $\mathrm{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\begin{aligned}
\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu = \int -f \log\left(\frac{g}{f}\right) \mathrm{d}\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \mathrm{d}\mu \\
= 1 + 1 - 2 \int \sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\
= \int f + g - 2\sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\
= \int (\sqrt{f} - \sqrt{g})^2\, \mathrm{d}\mu = d_{H^2}(f, g)
\end{aligned}$$

Showing $\mathrm{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\begin{aligned}
\int f \log\left(\frac{f}{g}\right) \mathrm{d}\mu &= \int -f \log\left(\frac{g}{f}\right) \mathrm{d}\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \mathrm{d}\mu \\
&= 1 + 1 - 2 \int \sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\
&= \int f + g - 2\sqrt{f}\sqrt{g}\, \mathrm{d}\mu \\
&= \int (\sqrt{f} - \sqrt{g})^2\, \mathrm{d}\mu = d_{H^2}(f, g)
\end{aligned}$$

Conclude:

$\mathrm{KL}(f, g) = 0 \implies d_{H^2}(f, g) = 0 \implies f = g \;\; \mu$–a.e.

Conclusion

It has now been shown that

$$\mathrm{KL}(f, g) \geq 0$$

Conclusion

It has now been shown that

$$\mathrm{KL}(f, g) \geq 0$$

$$\mathrm{KL}(f, g) = 0 \iff f = g \ \mu\text{–a.e.}$$

Conclusion

It has now been shown that

$\mathrm{KL}(f, g) \geq 0$

$\mathrm{KL}(f, g) = 0 \iff f = g$ μ–a.e.

Together this implies that $\mathrm{KL}(f, g)$ has a (μ–a.e.) unique minimum.

Conclusion

It has now been shown that

$\mathrm{KL}(f, g) \geq 0$

$\mathrm{KL}(f, g) = 0 \iff f = g$ μ–a.e.

Together this implies that $\mathrm{KL}(f, g)$ has a (μ–a.e.) unique minimum.

Thank you for your attention!