1. L(λ | xn) = λe^(-λx1)λe^(-λx2) … λe^(-λxn) =>

λ^n[e^(-λx1)e^(-λx2) … e^(-λxn)] =>

(d/dλ)log( λ^n) + log([e^-λ(x1 + x2 + … + xn)]) =>

(d/dλ)nlog(λ) - λ(x1 + x2 + … + xn) =>

0 = n(1/λ) - (x1 + x2 + … + xn) =>

λ = n/(x1 + x2 + … + xn)

1. nlog(λ) - λ(x1 + x2 + … + xn) + logP(λ) =>

0 = n(1/λ) - (x1 + x2 + … + xn) + 1/λ

𝜽𝑴𝑨𝑷 = n/(x1 + x2 + … + xn) + 1

logP(λ|D)∝logP(D|λ) +logP(λ)

1. P(D | λ)P(λ) = (λ^(Nxbar)e^-Nλ) \* (λ^(a-1)e^(-Bλ)) =>

λ^(Nxbar + a - 1)e^(-(B + N)λ) = gamma(Nxbar + a, B + N)

1. More features means more possible similarities in knn. The more data points means more possible weight to a Euclidean distance.

More lambda is less variance.

1. 1. 0.323625 = 32% is >50k
   2. With so few >50k points, 70% would not be very good for it as it has a high chance of missing that small group
   3. 84
2. △v = x - z => ||△v|| = ||x - z||=>the distance between two vectors in a straight line=>

L2 norm of the ||x-z||2

9.

Debriefing