

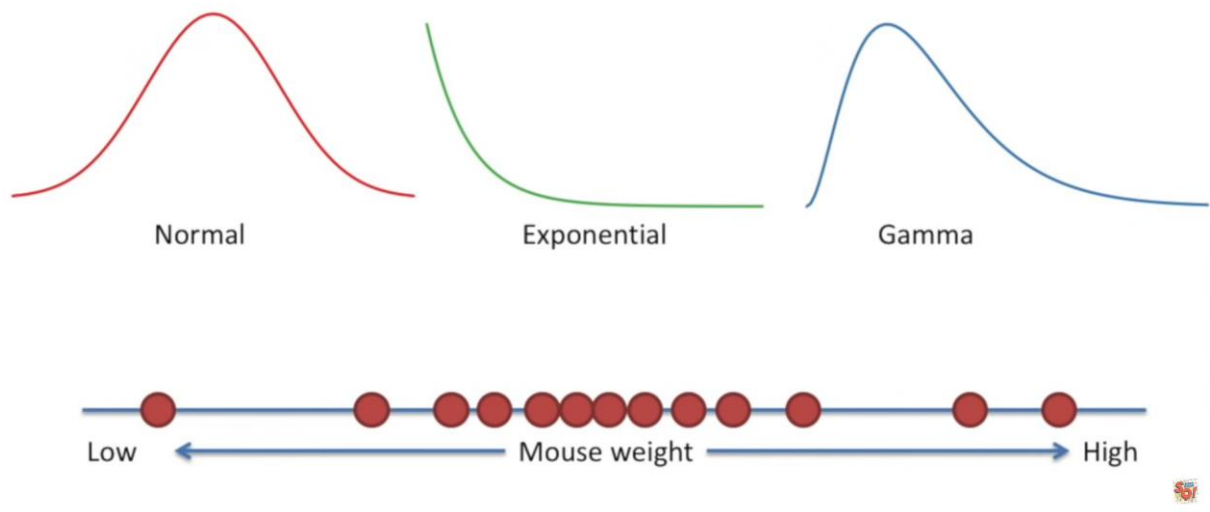
Maximum Likelihood

Maximum Likelihood

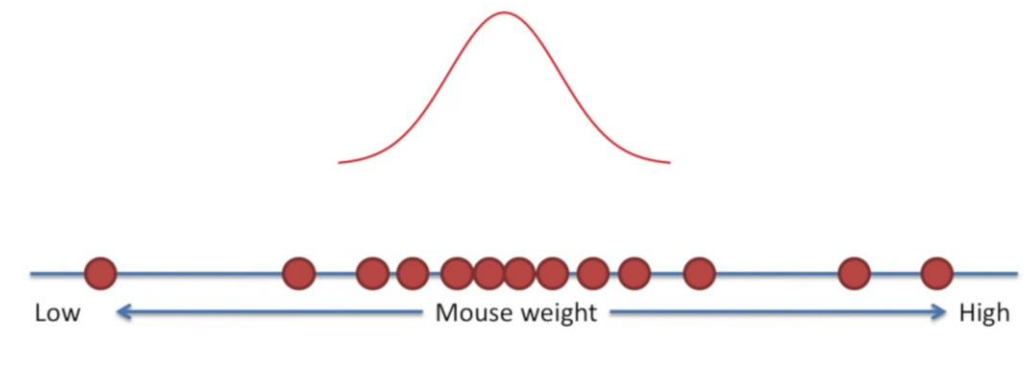
Diyelim ki bir grup fareyi tarttık. Maximum likelihood (maksimum olabilirliğin) amacı, bir dağılımı verilere uydurmanın en uygun yolunu bulmaktır.



Farklı veri türleri için *normal dağılım*, *exponential dağılım*, *gamma dağılımı* gibi birçok farklı dağılım türü vardır. Verilere bir dağılım uydurmak (fit etmek) istememizin nedeni, daha kolay çalışabilmesi ve daha genel olmasıdır - aynı türdeki her deney için geçerlidir.

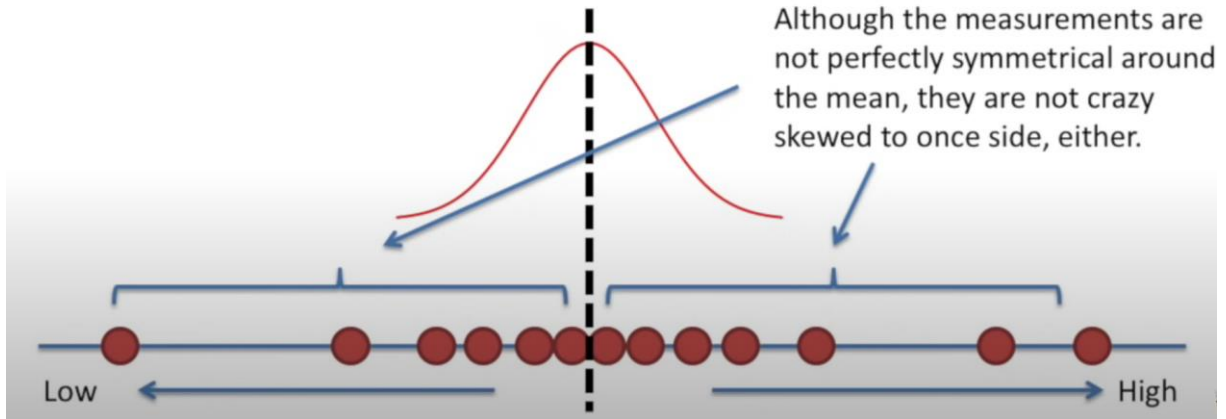


Bu case’te ağırlıkların normal dağılabileceğini düşünüyoruz. Bu, şu tür bir dağılımdan geldiğini düşündüğümüz anlamına gelir:

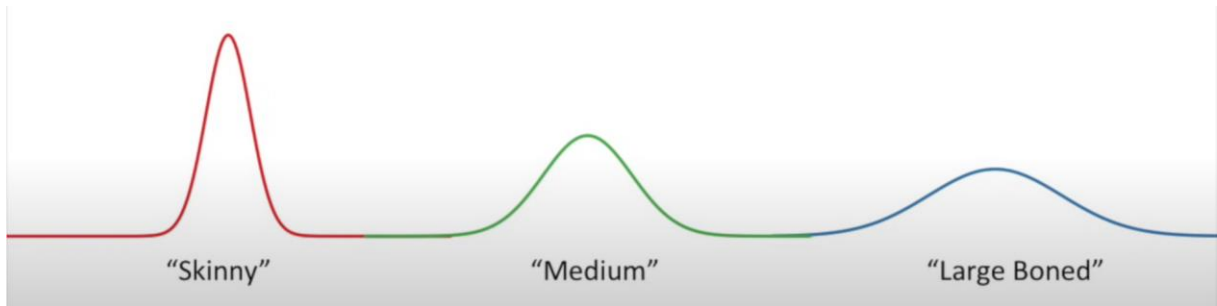


Normal dağılım şu anlamlara gelir:

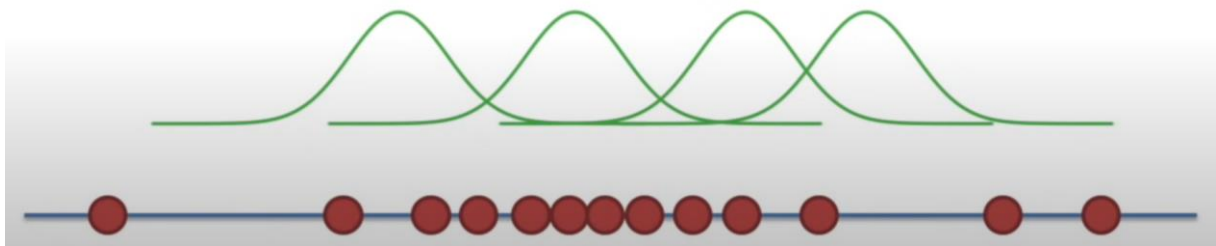
- 1) Ölçümlerin çoğunun (fare ağırlıkları) ortalamaya yakın olmasını bekleriz. Kendi data setimizde farelerin çoğunun ortalamaya yakın olduğunu görüyoruz.
- 2) Ölçümlerin ortalama etrafında nispeten simetrik olmasını bekliyoruz. Data setimizde, ölçümler ortalama etrafında tam simetrik olmasa da, bir tarafa çılgınca çarpık da değiller.



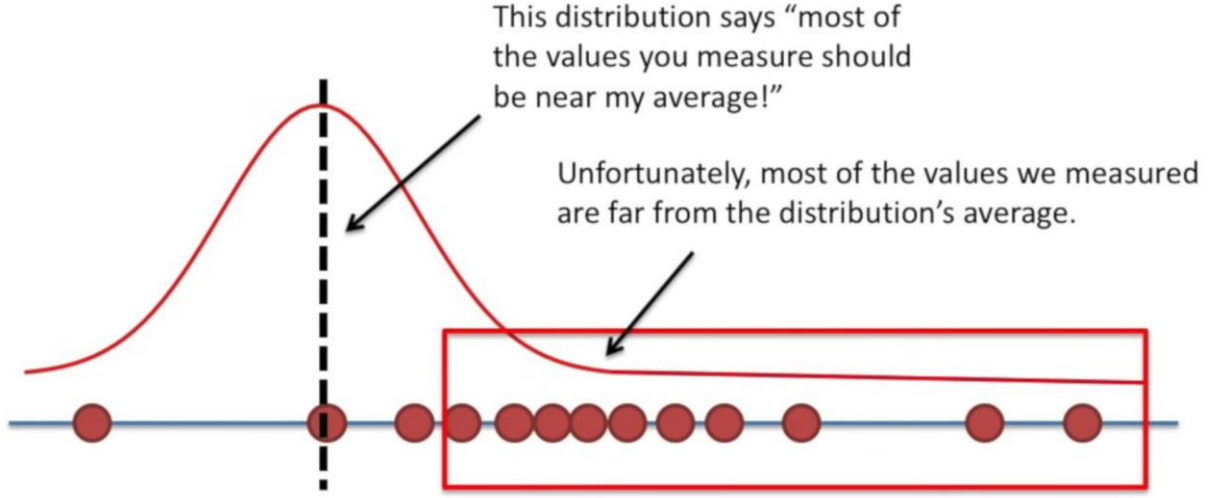
Normal dağılımlar her türlü şekil ve boyutta gelir. Skinny, medium veya large boned olabilirler.



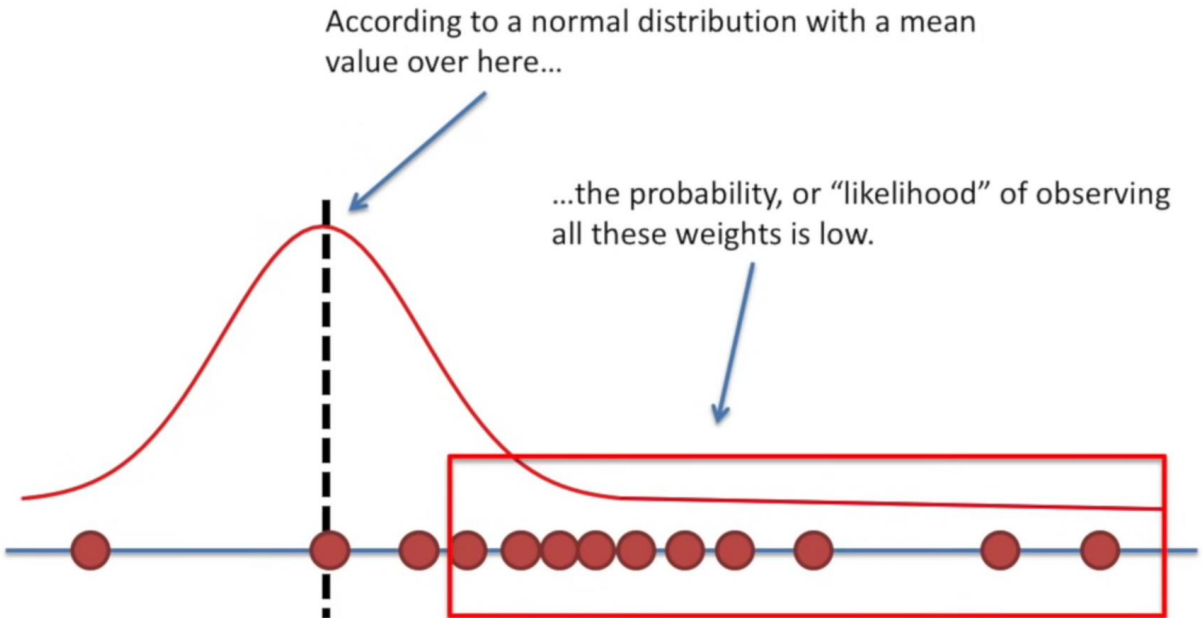
Şekle karar verdikten sonra, dağılımı nerede merkezleyeceğimizi bulmalıyız. Bir konum değerinden "daha iyi" midir?



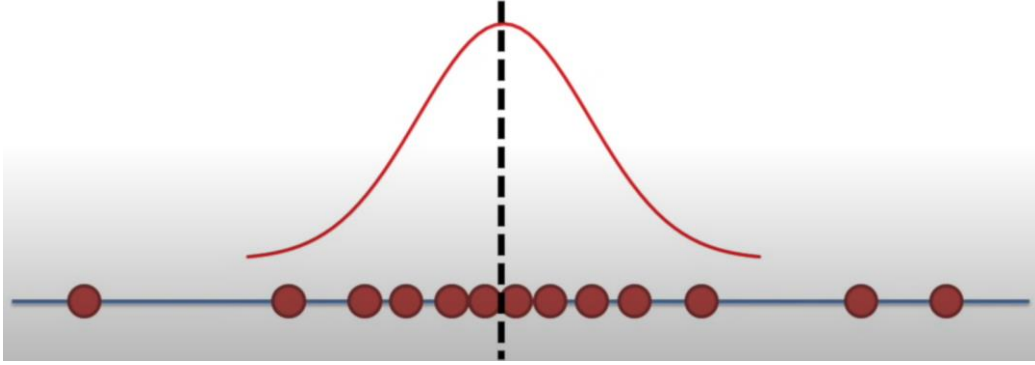
Tekniğe dalmadan önce, herhangi bir eski normal dağılımı seçelim ve verilere ne kadar uyduğunu görelim. Bu dağılım "ölçtüğünüz değerlerin çoğu benim ortalamama yakın olmalı!" diyor. Bu case'te, ortalama gerçek ölçümlerin ortalamasından farklıdır. Ne yazık ki ölçtüğümüz değerlerin çoğu dağılımın ortalamasından uzak.



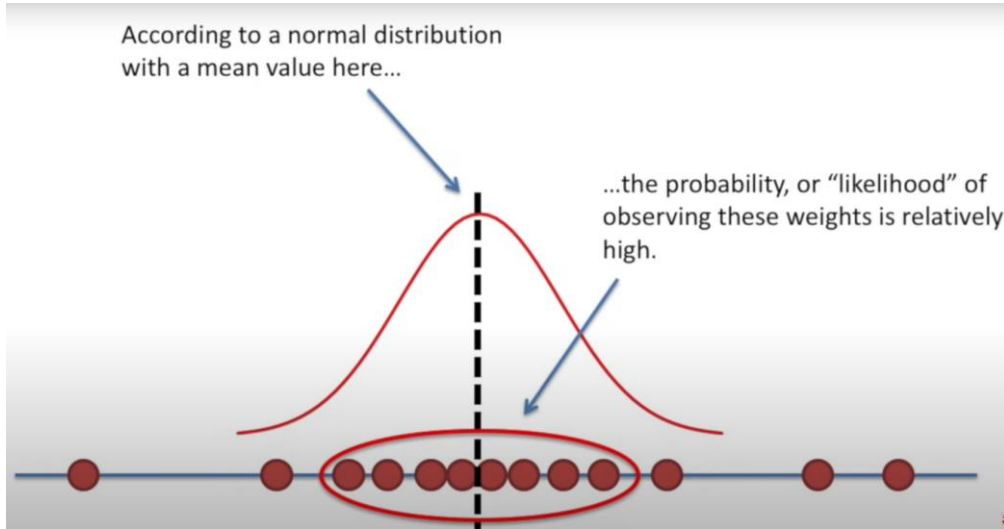
Bu ortalama değere sahip bir normal dağılıma göre gözlemlenen ağırlıkların probabilitysi veya likelihood düşük.



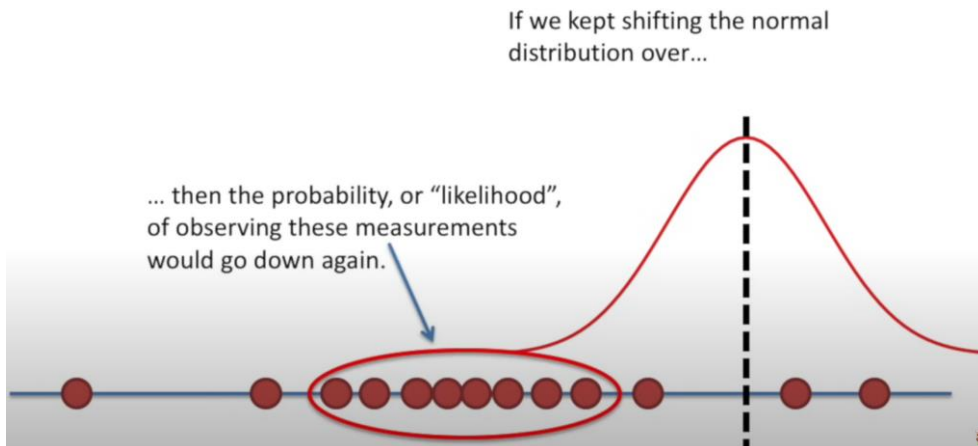
Normal dağılımı, ortalaması ortalama ağırlıkla aynı olacak şekilde kaydırırsak ne olur?



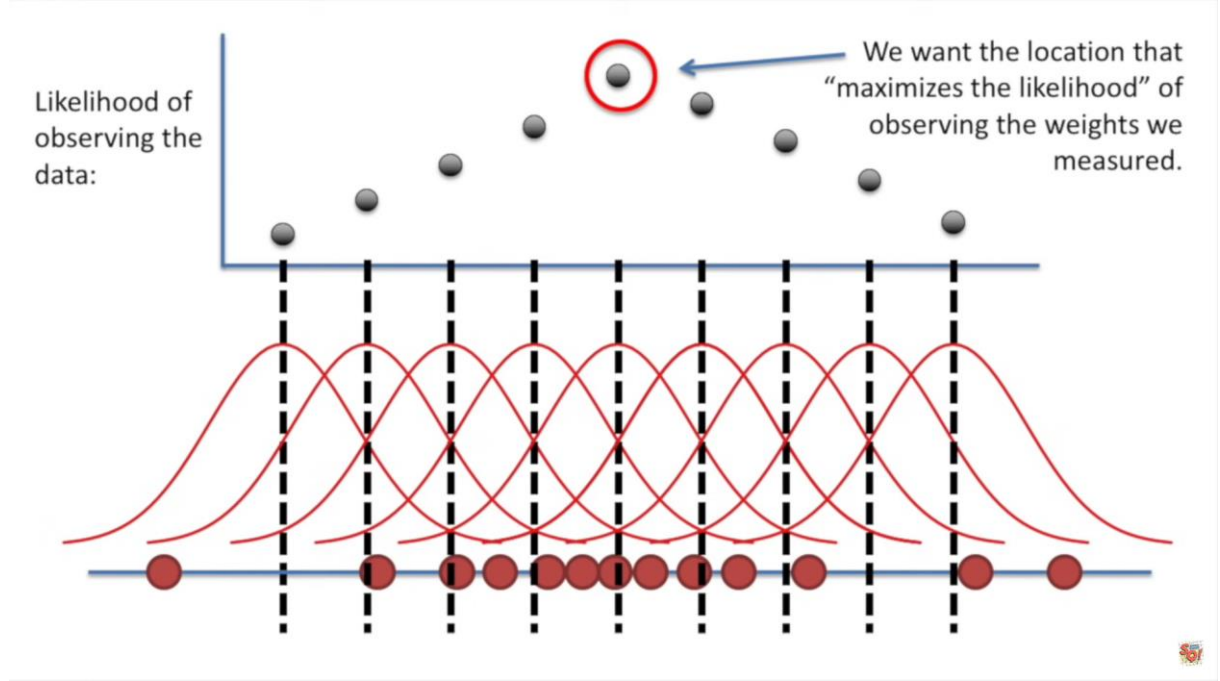
Ortalaması burada olan normal bir dağılıma göre, bu ağırlıkları gözlemleme probabilitysi veya likelihoodu nispeten yüksektir.



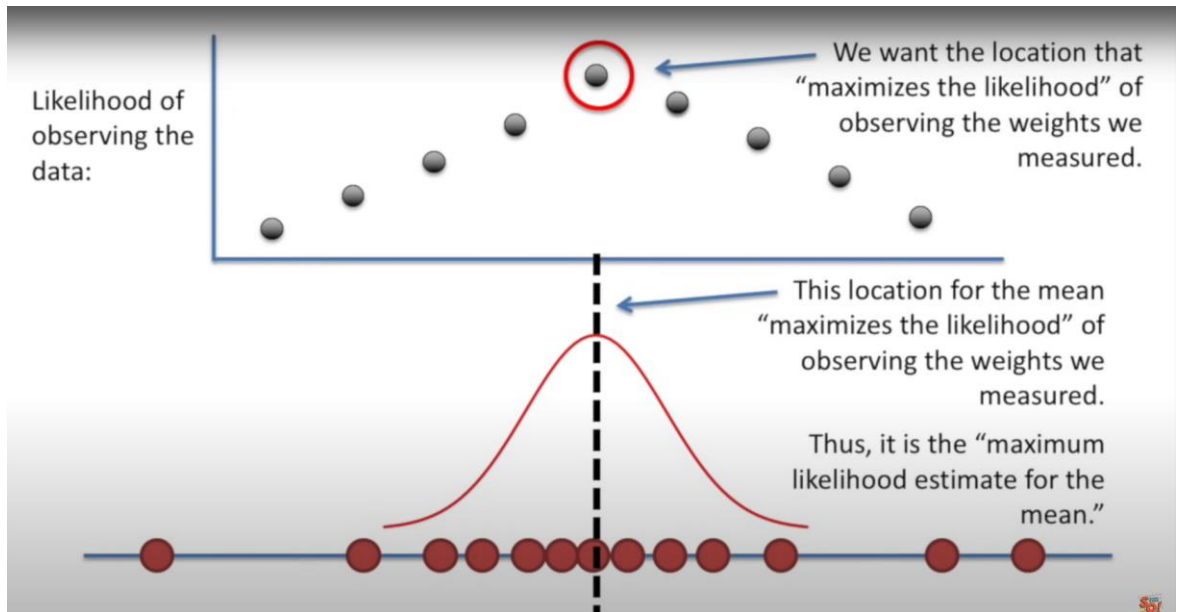
Normal dağılımı değiştirmeye devam edersek, bu ölçümleri gözlemleme probabilitysi veya likelihoodu tekrar düşecektir.



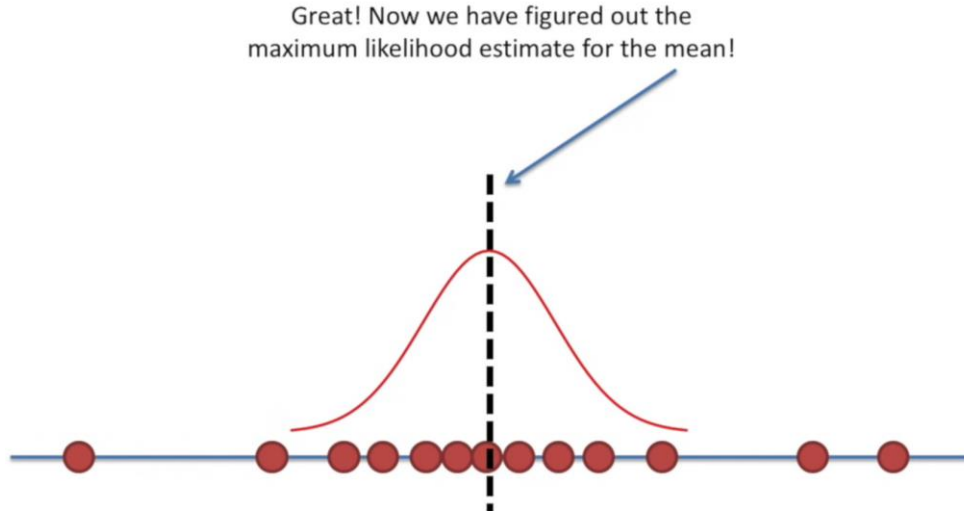
Verileri, dağılımın merkezinin konumu üzerinden gözlemleme likelihoodunu çizebiliriz. Sol taraftan başlıyoruz ve verileri gözlemleme olasılığını hesaplıyoruz ve ardından dağılımı sağa kaydırıp yeniden hesaplıyoruz. Bunu sadece verilerin sonuna kadar yapıyoruz. Tüm olası konumları denedikten sonra, normal dağılımı merkezleyebiliriz. Ölçtüğümüz ağırlıkları gözlemleme olasılığını (likelihood) en üst düzeye çıkaran konumu isteriz.



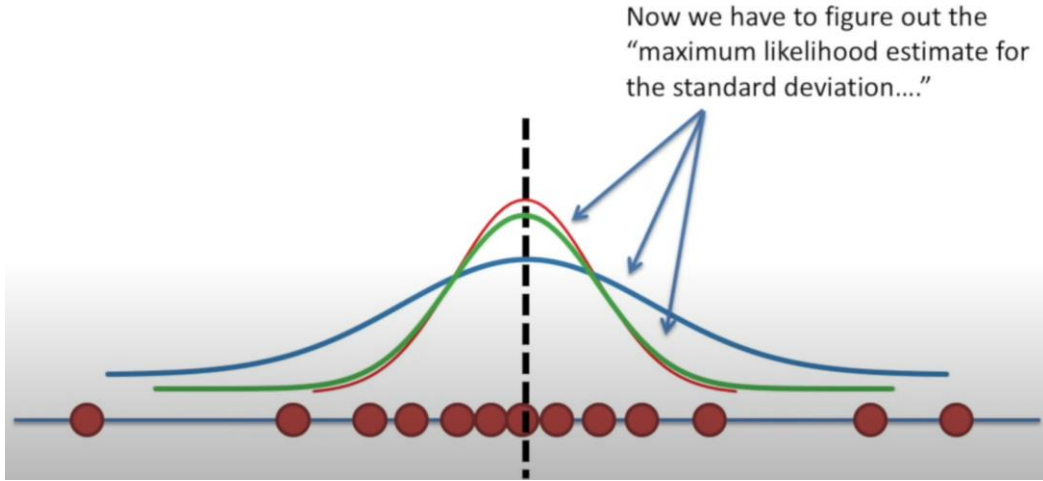
Ortalamanın bu konumu, ölçtüğümüz ağırlıkları gözlemleme olasılığını (likelihood) "maksimize eder". Bu nedenle, "ortalama için maksimum likelihood tahmini"dir. Bu durumda, verilerin ortalamasından değil, özellikle dağılımın ortalamasından bahsediyoruz. Ancak normal bir dağılımda bu iki şey aynıdır.



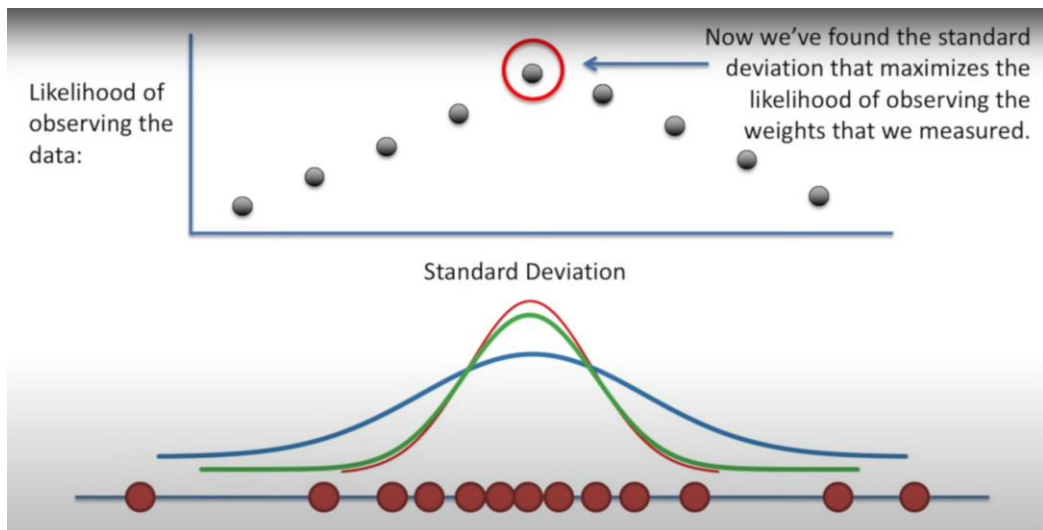
Ortalama için maksimum olabilirlik tahminini bulmuş olduk.



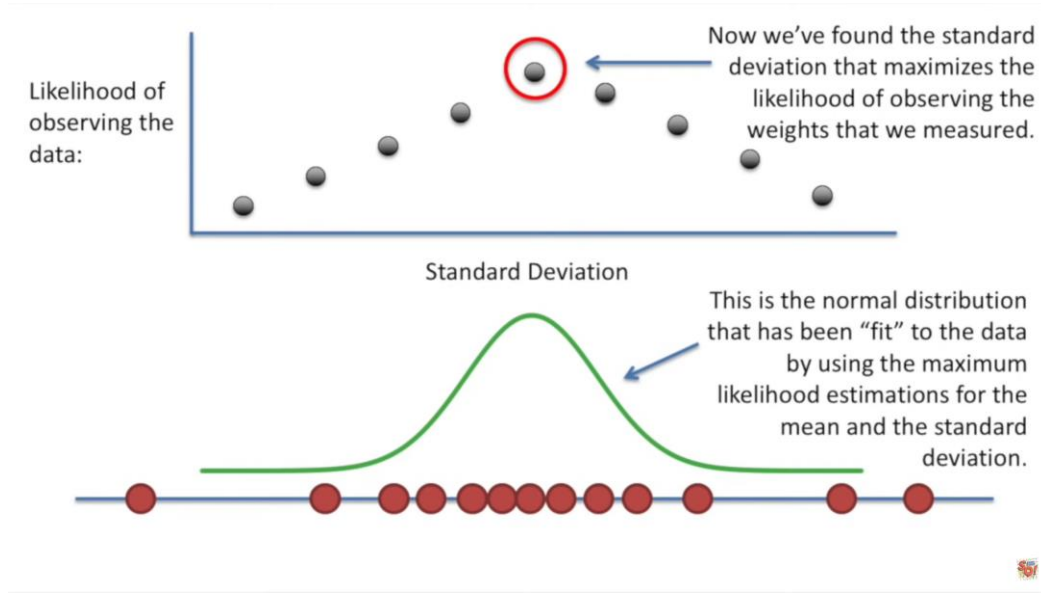
Şimdi standart sapma için maksimum likelihood tahminini bulmamız gerekiyor.



Yine, standart sapma için verileri farklı değerler üzerinden gözlemleme olasılığını çizebiliriz. Şimdi ölçtüğümüz ağırlıkları gözlemleme olasılığını maksimize eden standart sapmayı bulduk.



Bu, ortalama ve standart sapma için maksimum likelihood tahmini kullanılarak verilere uyan (fit) normal dağılımdır.



Şimdi birisi ortalama veya standart sapma veya başka bir şey için maksimum likelihood tahminlerine sahip olduğunu söylediğinde, gözlemlediğiniz şeyleri gözlemleme olasılığını (likelihood) en üst düzeye çıkaran ortalama veya standart sapma değerini bulduğunu bilirsiniz.

Not: Günlük konuşmada "probability" ve "likelihood" aynı anlama gelir. Ancak, Stats-Land'de "likelihood" özellikle burada ele aldığımız bu duruma atıfta bulunur; burada bir grup gözlenmiş ölçümü verilen dağılım için ortalama veya standart sapma için en uygun değeri bulmaya çalışıyoruz. Bir dağılımı verilere bu şekilde uydururuz.