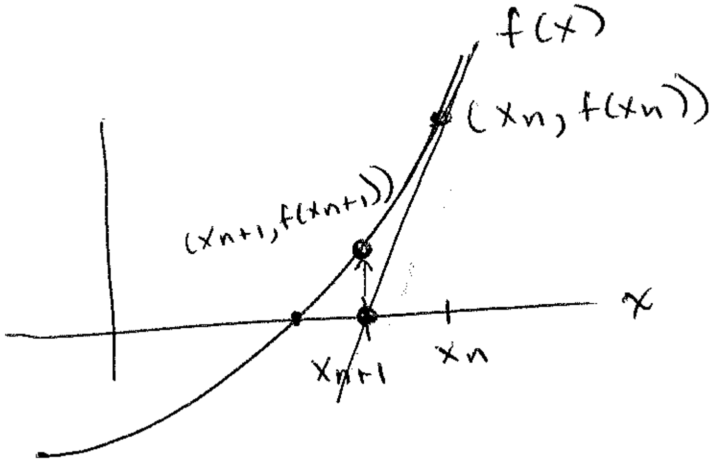
Geometric Intuition
Newton - Raphson



N - R is an iterative algo. that finds the <u>root</u> of a function.

Derive N-R:

Recall a line's slope: $\dfrac{y_2 - y_1}{x_2 - x_1}$

$$\frac{f(x_n) - 0}{x_n - x_{n+1}} = f'(x_n)$$

"slope of tangent line"

Solve for x_{n+1}:

$$f(x_n) = f'(x_n)(x_n - x_{n+1}) = f'(x_n)x_n - f'(x_n)x_{n+1}$$

$$f'(x_n)\,x_{n+1} = f'(x_n)\,x_n - f(x_n)$$

$$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}$$

Our example of a "root-finding" Scenario:

Find $\hat{\beta}_{MLE}$.

Let $\ell(\beta)$ be the log-likelihood function for some univariate β.

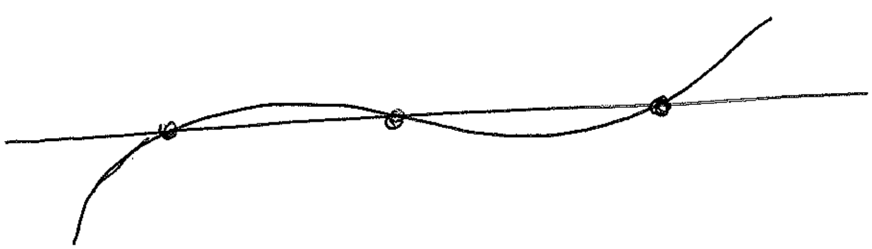
We want $\ell'(\hat{\beta}) = 0$ (MLE)

$\hookrightarrow$ this is like "f" in N-R.

N-R: $$\beta_{n+1} = \beta_n - \frac{\ell'(\beta_n)}{\ell''(\beta_n)}$$

Question in class: does the starting pt matter?

Answer: <u>yes</u>, e.g. imagine if objective function has multiple roots:



where we start can impact the root we find.

Logistic Regression Assumptions

$$p(Y_i = 1) = p_i = f(\underline{\beta})$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \underline{x}_i^T \underline{\beta}$$

Assumption:

1) $p_i \neq 0$ & $p_i \neq 1$
2) log odds have a linear relationship w/ predictors.

In writing likelihood:

$$p(\underline{y} \mid \underline{p}) = \prod_{i=1}^{n} p(y_i \mid p_i)$$

Assumption:

Outcomes are conditionally independent

$$\log p(\underline{y} \mid \underline{p}) = \sum_{i=1}^{n} \log(p(y_i \mid p_i))$$

BUILDING INTUITION:

Why χ^2?
Why -2?
(sketch Wilk's thm)

$\ell(\underline{\beta}) :=$ log-likelihood function of $\underline{\beta}$.

Idea: Taylor expand about $\underline{\hat{\beta}_n} \leftarrow$ MLE. [notice: dependence on # samples n.]

$\cancel{\ell(\hat{\beta}_n) = \ell(\hat{\beta}_n)}$

$$\ell(\underline{\beta}) = \ell(\hat{\beta}_n) + \underbrace{\cancel{\frac{\ell'(\underline{\hat{\beta}_n})(\underline{\beta}-\underline{\hat{\beta}_n})}{1!}}}_{0} + \frac{(\underline{\beta}-\underline{\hat{\beta}_n})^T \overbrace{\ell''(\underline{\hat{\beta}_n})}^{\text{Hessian matrix}} (\underline{\beta}-\underline{\hat{\beta}_n})}{2!} + \underbrace{\text{HOTS}}_{\text{higher order terms.}}$$

Remember $\underline{\hat{\beta}_n}$ is the value such that $\ell'(\hat{\beta}_n)=0$

I'm left w/ the following approximation:

$$\ell(\underline{\beta}) \approx \ell(\underline{\hat{\beta}_n}) + \frac{(\underline{\beta}-\underline{\hat{\beta}_n})^T \ell''(\underline{\hat{\beta}_n})(\underline{\beta}-\underline{\hat{\beta}_n})}{2}$$

move over

$\Rightarrow$

$$\boxed{2\left(\ell(\underline{\beta}) - \ell(\underline{\hat{\beta}_n})\right) \approx (\underline{\beta}-\underline{\hat{\beta}_n})^T \ell''(\underline{\hat{\beta}_n})(\underline{\beta}-\underline{\hat{\beta}_n})}$$

Questions: (1) how is this a likelihood ratio?
(2) how is this χ^2?

Answers (1): property of <u>log</u>. $\log(a)-\log(b) = \log(a/b)$
(2) $\underline{\hat{\beta}_n}$ is MLE. Therefore asymptotically <u>normal</u> w/ mean $\underline{\beta}$ (true popln parameter) & variance $\ell''(\underline{\beta}_n)^{-1}$

This means the term

$$(\underline{\beta} - \hat{\underline{\beta}}_n)^T \, l''(\hat{\underline{\beta}}_n) \, (\underline{\beta} - \underline{\beta}_n) \quad \text{is}$$

the sum of standard normals squared for large n. AUA, χ^2.