1. Linear regression models, using lm() function in R:

$Y_i = x_i^T \beta + \varepsilon_i$, ε_i iid $N(0, \sigma^2)$.

$E(Y_i) = x_i^T \beta \in \mathbb{R}$

$x_i^T \beta$:
1) one or more quantitative covariates
2) one or more categorical covariates
3) polynomial regression, in one or more variables.
4) models with interactions
5) all of the above.

2. GLM: "G" - generalized; useful when there are natural restrictions on ~~the~~ $E(Y_i)$.

Model $h(E(Y_i))$; $h: (0,1) \mapsto \mathbb{R}$ for our classification models.

Want: h must be strictly increasing.

E.g., let $Y_i \sim \mathrm{Bern}(p_i = p(x_i; \beta))$; Y_i are indep.

$E(Y_i) = 1 \cdot p_i + 0 \cdot (1-p_i) = p_i \in (0,1)$.

E.g., let F be the cdf of a cts rv; strictly increasing on $\mathbb{R}$.
take h to be the inverse F. ~~Let~~

E.g. 1: $F(z) = e^z/(1+e^z)$: logistic cdf. then $h(p) = F^{-1}(p) = \log \frac{p}{1-p}$

E.g. 2: $F(z) = \Phi(z)$: cdf of $N(0,1)$. $\Rightarrow$ probit regression

②

```
> glm.fits=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume ,
    data=Smarket ,family=binomial)
> summary(glm.fits)

Call:
glm(formula = Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5
    + Volume , family = binomial, data = Smarket)

Deviance Residuals:
   Min      1Q   Median      3Q     Max
 -1.45   -1.20    1.07    1.15    1.33

Coefficients:
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept) -0.12600     0.24074    -0.52      0.60
Lag1        -0.07307     0.05017    -1.46      0.15
Lag2        -0.04230     0.05009    -0.84      0.40
Lag3         0.01109     0.04994     0.22      0.82
Lag4         0.00936     0.04997     0.19      0.85
Lag5         0.01031     0.04951     0.21      0.83
Volume       0.13544     0.15836     0.86      0.39
```

$Y_i \overset{indep}{\sim} \mathrm{Bern}(p_i)$

$p_i = E(Y_i)$

$= \mathrm{expit}(x_i^T \beta)$.

inverse of the logit transform.

$\hat{\beta}$

square roots of diag ($\hat{\Sigma}$)

$\dfrac{\hat{\beta}_1}{\sqrt{\hat{\Sigma}_{22}}}$

$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_6 \end{pmatrix}$

: pvalue based on std. normal cdf.

MLE estimation: ① $\mathrm{Lik}(\beta) = \prod_{i=1}^{n} p_i^{y_i} \cdot (1-p_i)^{(1-y_i)}$.

② $\log.\mathrm{lik}(\beta) = \sum_{i=1}^{n} y_i \ln p_i + (1-y_i) \cdot \ln(1-p_i)$.

③. $g(\beta) = -\log.\mathrm{lik}(\beta)$. $\Rightarrow$ minimize g numerically; $\hat{\beta} = \arg\min_{\beta} g(\beta)$.

$\hat{H} = \nabla^2_{\beta\beta}\, g(\beta)\big|_{\beta=\hat{\beta}}$: Hessian. Let $\hat{\Sigma} = (\hat{H})^{-1}$. Asympt theory suggests $\hat{\beta} \overset{D}{\approx} \mathrm{MVN}(\beta_{true}, \hat{\Sigma})$.