Things to Memorize

- Defn. of likelihood (Bernoulli):  \( P(\mathcal{D} \mid \theta) = \theta^{\alpha_H}(1 - \theta)^{\alpha_T} \)
- Beta prior:  
  \[
P(\theta) = \frac{\theta^{\beta_H-1}(1 - \theta)^{\beta_T-1}}{B(\beta_H, \beta_T)} \sim \text{Beta}(\beta_H, \beta_T)
\]
  (You can ignore the normalization constant and just remember the proportionality.)
- MLE:  
  \[
  \hat{\theta}_{\text{MLE}} = \frac{\alpha_H}{\alpha_H + \alpha_T}
  \]
- MAP:  
  \[
  \hat{\theta} = \arg \max_{\theta} P(\theta \mid \mathcal{D}) = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}
  \]
- Perceptron update:  
  \[
  w_i \leftarrow w_i + \eta(t - o)x_i
  \]
- Entropy:  
  \[
  H(V) = -P(V = 1)\log P(V = 1) - P(V = 0)\log P(V = 0)
  \]
- Information gain:
  \[
  G(V,A) = H(V) - P(A = 0)H(V \mid A = 0) - P(A = 1)H(V \mid A = 1)
  \]
  where \( H \) is entropy and \( H(V \mid A=0) \) indicates the entropy computed over only the examples where \( A=0 \).