

Lab 4 — Baum HMM Regime Detector

David K.H. Chan · April 2026

Objective

Build a regime detector from first principles using Baum's Hidden Markov Model. The output is not a hard label — it's a probability distribution over hidden states at every bar. These soft posteriors $\gamma_k(i)$ feed directly into Labs 5, 6, 7, and 8.

Model Architecture

Parameter	Value
Asset	BTC-USD daily candles (2017–2026)
Hidden states	3 — Bull, Choppy, Bear
Observation features	fracChange, fracHigh, fracLow
Algorithm	Baum-Welch EM (Gaussian HMM)
Restarts	10 random initialisations, best log-likelihood selected
Library	hmmlearn

Observation Vector

Three scale-invariant candle shape features — identical across any price level:

```
fracChange = (close - open) / open    ← body direction and size
fracHigh   = (high  - open) / open    ← upper wick
fracLow    = (open  - low)  / open    ← lower wick
```

These capture candle *shape*, not absolute price. A \$100 candle and a \$100,000 candle with the same shape produce identical observations.

Algorithm — Baum-Welch (Forward-Backward)

Baum-Welch is an EM algorithm for HMMs. At each iteration:

- **E-step (Forward-Backward):** Compute $\gamma_k(i) = P(\text{state}=k \text{ at time } i \mid \text{all observations})$
- **M-step:** Re-estimate emission means/covariances and transition matrix to maximise log-likelihood

The output at every bar is a probability distribution over states — not a hard label. This is the key distinction from simple threshold-based regime detection.

State Labelling

States are labelled automatically by sorting on mean `fracChange`:

- **Highest mean return** → Bull
- **Middle** → Choppy
- **Lowest mean return** → Bear

No manual tuning. The model discovers the structure from data.

Transition Matrix

The learned transition matrix encodes regime persistence — how likely each regime is to continue vs. switch. Bull and Bear regimes are typically sticky (diagonal entries > 0.9), while Choppy is the transitional state.

Output for Downstream Labs

Lab 4 exports `lab4_output.csv` with the full posterior matrix:

```
gamma_0, gamma_1, gamma_2  ← soft posteriors at every bar
state                        ← Viterbi hard label
regime                      ← "Bull" / "Choppy" / "Bear"
```

These columns are consumed directly by: - **Lab 5** — XGBoost features (γ_k as inputs)
- **Lab 6** — SOL-specific HMM retraining - **Lab 8** — Position sizing via regime conviction

Key Insight

Baum (1970) independently discovered the same forward-backward recurrence that appears in Kolmogorov's equations and modern deep learning backpropagation. The HMM is a concrete implementation of a universal mathematical structure for inferring hidden state from noisy observations.

The $\gamma_k(i)$ posteriors are the foundation of the entire Fortuna stack.

Connection to Fear & Greed

A simple alternative to the HMM is the Crypto Fear & Greed Index (Alternative.me API). It returns a daily 0–100 score that roughly maps to:

Score	Regime
0–25	Bear
25–45	Choppy
45–75	SoftBull
75–100	Bull

The HMM is more precise (it learns asset-specific regime boundaries from price structure) but the Fear & Greed index is a useful sanity check. When both agree → high confidence. When they disagree → flag the signal.