https://github.com/ddbourgin/numpy-ml/tree/master

1. **Gaussian mixture model**
   * EM training
2. **Hidden Markov model**
   * Viterbi decoding
   * Likelihood computation
   * MLE parameter estimation via Baum-Welch/forward-backward algorithm
3. **Latent Dirichlet allocation** (topic model)
   * Standard model with MLE parameter estimation via variational EM
   * Smoothed model with MAP parameter estimation via MCMC
4. **Neural networks**
   * Layers / Layer-wise ops
     + Add
     + Flatten
     + Multiply
     + Softmax
     + Fully-connected/Dense
     + Sparse evolutionary connections
     + LSTM
     + Elman-style RNN
     + Max + average pooling
     + Dot-product attention
     + Embedding layer
     + Restricted Boltzmann machine (w. CD-n training)
     + 2D deconvolution (w. padding and stride)
     + 2D convolution (w. padding, dilation, and stride)
     + 1D convolution (w. padding, dilation, stride, and causality)
   * Modules
     + Bidirectional LSTM
     + ResNet-style residual blocks (identity and convolution)
     + WaveNet-style residual blocks with dilated causal convolutions
     + Transformer-style multi-headed scaled dot product attention
   * Regularizers
     + Dropout
   * Normalization
     + Batch normalization (spatial and temporal)
     + Layer normalization (spatial and temporal)
   * Optimizers
     + SGD w/ momentum
     + AdaGrad
     + RMSProp
     + Adam
   * Learning Rate Schedulers
     + Constant
     + Exponential
     + Noam/Transformer
     + Dlib scheduler
   * Weight Initializers
     + Glorot/Xavier uniform and normal
     + He/Kaiming uniform and normal
     + Standard and truncated normal
   * Losses
     + Cross entropy
     + Squared error
     + Bernoulli VAE loss
     + Wasserstein loss with gradient penalty
     + Noise contrastive estimation loss
   * Activations
     + ReLU
     + Tanh
     + Affine
     + Sigmoid
     + Leaky ReLU
     + ELU
     + SELU
     + GELU
     + Exponential
     + Hard Sigmoid
     + Softplus
   * Models
     + Bernoulli variational autoencoder
     + Wasserstein GAN with gradient penalty
     + word2vec encoder with skip-gram and CBOW architectures
   * Utilities
     + col2im (MATLAB port)
     + im2col (MATLAB port)
     + conv1D
     + conv2D
     + deconv2D
     + minibatch
5. **Tree-based models**
   * Decision trees (CART)
   * [Bagging] Random forests
   * [Boosting] Gradient-boosted decision trees
6. **Linear models**
   * Ridge regression
   * Logistic regression
   * Ordinary least squares
   * Weighted linear regression
   * Generalized linear model (log, logit, and identity link)
   * Gaussian naive Bayes classifier
   * Bayesian linear regression w/ conjugate priors
     + Unknown mean, known variance (Gaussian prior)
     + Unknown mean, unknown variance (Normal-Gamma / Normal-Inverse-Wishart prior)
7. **n-Gram sequence models**
   * Maximum likelihood scores
   * Additive/Lidstone smoothing
   * Simple Good-Turing smoothing
8. **Multi-armed bandit models**
   * UCB1
   * LinUCB
   * Epsilon-greedy
   * Thompson sampling w/ conjugate priors
     + Beta-Bernoulli sampler
   * LinUCB
9. **Reinforcement learning models**
   * Cross-entropy method agent
   * First visit on-policy Monte Carlo agent
   * Weighted incremental importance sampling Monte Carlo agent
   * Expected SARSA agent
   * TD-0 Q-learning agent
   * Dyna-Q / Dyna-Q+ with prioritized sweeping
10. **Nonparameteric models**
    * Nadaraya-Watson kernel regression
    * k-Nearest neighbors classification and regression
    * Gaussian process regression
11. **Matrix factorization**
    * Regularized alternating least-squares
    * Non-negative matrix factorization
12. **Preprocessing**
    * Discrete Fourier transform (1D signals)
    * Discrete cosine transform (type-II) (1D signals)
    * Bilinear interpolation (2D signals)
    * Nearest neighbor interpolation (1D and 2D signals)
    * Autocorrelation (1D signals)
    * Signal windowing
    * Text tokenization
    * Feature hashing
    * Feature standardization
    * One-hot encoding / decoding
    * Huffman coding / decoding
    * Byte pair encoding / decoding
    * Term frequency-inverse document frequency (TF-IDF) encoding
    * MFCC encoding
13. **Utilities**
    * Similarity kernels
    * Distance metrics
    * Priority queue
    * Ball tree
    * Discrete sampler
    * Graph processing and generators