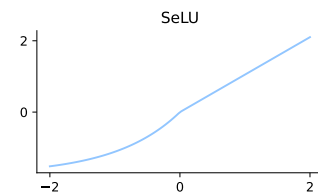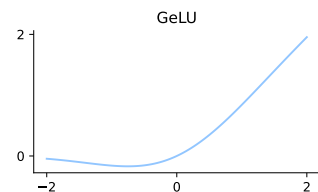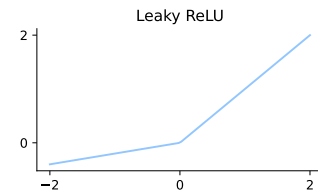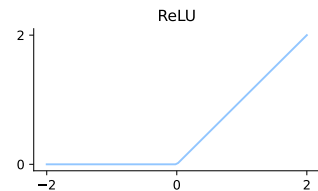# A Practical Guide to Deep Network Design

Loss Functions
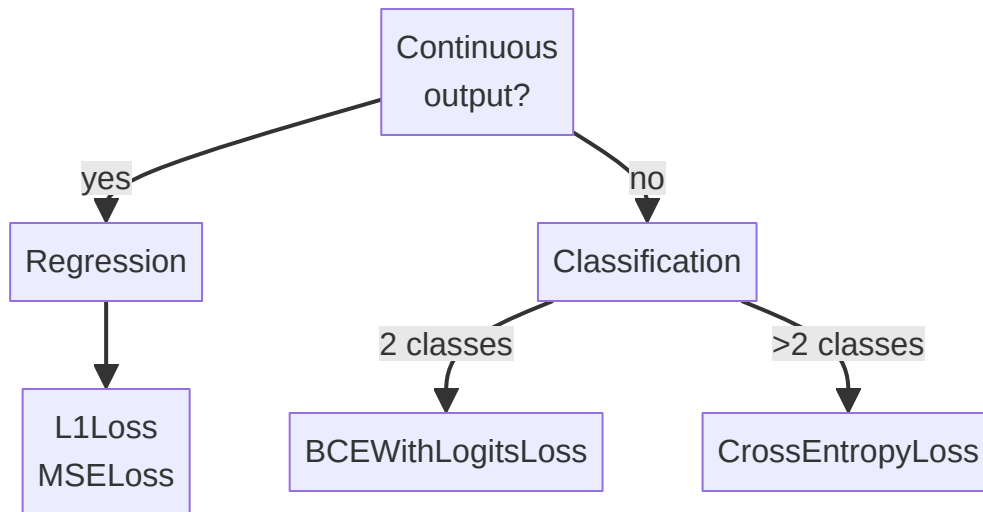
Activation Functions

- Try ReLU
- If ReLU fails, try:
    - Leaky ReLU / PReLU
- GeLU for sophisticated models

SGD

- Select largest possible batch size
- momentum = 0.9
- Tune learning rate

```
b = 0
for epoch in range(n):
  for i in range(len(dataset) // batch_size):
    batch = dataset[i * batch_size : (i + 1) * batch_size]
    for (x, y) in batch:
      b = ∇l(θ|x,y) + momentum * b
      θ = θ - ε * b.mT
```