| **Task Type** | **# of Classes** | **Target Format** | **Last Layer Activation** | **Loss Function** | **Notes** |
| --- | --- | --- | --- | --- | --- |
| **Binary Classification** | 2 | 0 or 1 (float or long) | Sigmoid or *None* | BCELoss / BCEWithLogitsLoss | Use BCEWithLogitsLoss **without Sigmoid** for numerical stability |
| **Multiclass Classification** | >2 | Class index (e.g., 0, 1, 2) | *None* | CrossEntropyLoss | CrossEntropyLoss internally applies LogSoftmax, so no activation |
| **Multilabel Classification** | >2 (multi-label) | Multi-hot vector (e.g., [1,0,1]) | Sigmoid | BCELoss / BCEWithLogitsLoss | Each label is treated independently |
| **Regression (Single Output)** | 1 | Float (e.g., 3.75) | *None* | MSELoss / L1Loss | No activation needed, outputs raw values |
| **Regression (Multi Output)** | >1 | Float vector (e.g., [2.1, 3.5]) | *None* | MSELoss / L1Loss | Final Linear layer with output size = # of targets |
| **Ordinal Classification** | Ordered classes | Custom encoding | Custom | CrossEntropyLoss, OrdinalLoss | Requires task-specific logic or specialized loss |
| **Soft Classification / Probabilities** | >2 | Probabilities or float targets | Softmax | KLDivLoss or custom loss | Rare, use when targets are distributions rather than class labels |