These notes summarise selected lecture concepts and are not a substitute for working through the lecture slides, tutorials, and self-study exercises. Feel free to personalise and develop them into your own summary sheet.

Loss $\ell(\mathbf{h}, \mathbf{a})$ — We take action $\mathbf{a}$ without observing $\mathbf{h}$; once $\mathbf{h}$ is revealed, we incur loss $\ell(\mathbf{h}, \mathbf{a})$.

Risk $R(\mathbf{a}; f)$ — The expected loss $\mathbb{E}_{f(\mathbf{h})}[\ell(\mathbf{h}, \mathbf{a})]$, where f denotes the distribution over $\mathbf{h}$, potentially being a prior belief $p(\mathbf{h})$, a posterior $p(\mathbf{h}|\mathbf{x})$, or the data distribution.

Empirical risk $\hat{R}_n(\mathbf{a}; f)$ — The risk when the expectation over f is approximated with a sample average based on n samples $\mathbf{h}_i \sim f(\mathbf{h})$, $\hat{R}_n(\mathbf{a}; f) = \frac{1}{n}\sum_{i=1}^{n} \ell(\mathbf{h}_i, \mathbf{a})$.

Expected loss principle — We choose the action $\mathbf{a}^*(f)$ that minimises the (empirical) risk.

Minimax principle — We choose the action that minimises the worst-case loss $\max_{\mathbf{h}} \ell(\mathbf{h}, \mathbf{a})$.

Policy — A mapping from evidence to action.

Common loss functions and actions that minimise the risk—

loss	$\ell(\mathbf{h}, \mathbf{a})$	optimal action $\mathbf{a}^*(f)$
quadratic	$(h-a)^2$	$\mathbb{E}_{f(h)}[h]$
absolute error	$\lvert h-a\rvert$	median(f)
0-1	$1 - \mathbb{1}(\mathbf{h} = \mathbf{a})$	$\operatorname{argmax}_{\mathbf{h}} \log f(\mathbf{h})$
log$^{(*)}$	$-\log q(\mathbf{h})$	f

$^{(*)}$ For the log-loss, the action $\mathbf{a}$ is a distribution $q(\mathbf{h})$.

Cross-entropy — The risk under the log-loss, $\mathbb{E}_{f(\mathbf{h})}[-\log q(\mathbf{h})] = \mathrm{KL}(f||q) + \text{entropy of } f$. Minimised by $q = f$ since $\mathrm{KL}(f||q) \geq 0$ with equality iff $q = f$.

MAP estimate — Minimises the posterior expected 0-1 loss: $\hat{\boldsymbol{\theta}}_{MAP} = \operatorname{argmax}_{\boldsymbol{\theta}} \log p(\mathcal{D}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})$. The log-prior acts as a regulariser whose influence fades as more data are observed.

Posterior — Minimises the posterior expected log-loss: $p(\boldsymbol{\theta}|\mathcal{D}) = \operatorname{argmin}_q \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}[-\log q(\boldsymbol{\theta})]$.

Maximum likelihood estimation — Frequentist view: minimising empirical risk under log-loss over a parametric family $\mathcal{Q}$. Bayesian view: MAP estimate with a flat prior.