Example Language and Notation for ML Course

1. [projections and orthonormal vectors] Let S be the subspace spanned by the orthonormal vectors a and b. Let p be the projection of the vector v into S. Let $r = v - p$ be the residual vector. Then $r \perp S$ and $\{r, a, b\}$ form an orthonormal set.

2. [linear ridge regression] Given some data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbf{R}^d \times \mathbf{R}$, the ridge regression solution for regularization parameter $\lambda > 0$ is given by
$$\hat{w} = \underset{w \in \mathbf{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left\{ w^T x_i - y_i \right\}^2 + \lambda \|w\|_2^2,$$
where $\|w\|_2^2 = w_1^2 + \cdots + w_d^2$ is the square of the ℓ_2-norm.

3. [completing the square] You should be able to verify, just by multiplying out the expressions on the RHS, that the following "completing the square" identity is true: For any vectors $x, b \in \mathbf{R}^d$ and symmetric invertible matrix $M \in \mathbf{R}^{d \times d}$, we have
$$x^T M x - 2b^T x = \left(x - M^{-1}b\right)^T M (x - M^{-1}b) - b^T M^{-1} b \quad (0.1)$$

4. [taking a gradient] You should be comfortable taking the gradient of the following w.r.t. w:
$$L(w, b, \xi, \alpha, \lambda) = \frac{1}{2}||w||^2 + \frac{c}{n} \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \alpha_i \left(1 - y_i \left[w^T x_i + b\right] - \xi_i\right) - \sum_{i=1}^{n} \lambda_i \xi_i$$

5. [directional derivative] If we fix a direction $u \in \mathbf{R}^d$, we can compute the directional derivative $f'(x; u)$ as
$$f'(x; u) = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}.$$

6. [the risk functional] For "loss" function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbf{R}$, define the "risk" of a function $f : \mathcal{X} \to \mathcal{Y}$ by

$$R(f) = \mathbb{E}\ell\left(f(x), y\right),$$

where the expectation is over $(x, y) \sim P_{\mathcal{X}\times\mathcal{Y}}$, a distribution over $\mathcal{X} \times \mathcal{Y}$.

7. [unbiased estimate]. Consider $x_1, \ldots, x_n$ sampled i.i.d. from a distribution P on $\mathbf{R}$. Write $\mu = \mathbb{E}x$, for $x \sim P$. Then the mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an unbiased estimate of μ, since $\mathbb{E}\bar{x} = x$. Similarly, the sample variance $\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ is an unbiased estimate for $\mathrm{Var}\left(x\right)$. You should be able to easily verify these facts fairly easily.