DSGA-1003 Machine Learning and Computational Statistics
Prerequisite Questionnaire

1 Doing Data Science in Python

1. Rate your comfort level with Python:
   - ☐ Expert - I could (or do) get paid for it.
   - ☐ Good enough to get the job done.
   - ☐ Mmmm... Haven't used it much, but you know one language, you know them all, right?
   - ☐ Weird - why are you asking about snakes?

2. Rate your comfort level with numpy (http://www.numpy.org/):
   - ☐ I'm pretty proficient in numpy.
   - ☐ Not so much, but I'm good at matrix/vector stuff in matlab (or something else), and I'm very comfortable with vectorizing mathematical calculations.
   - ☐ Can't wait to learn!
   - ☐ Not super-excited about programming learning algorithms from scratch – hasn't somebody else already solved that problem for us?

3. Rate your fluency in data visualization in Python (e.g. matplotlib, bqplot, etc.)
   - ☐ I make great plots.
   - ☐ With enough googling, I can get the job done.
   - ☐ I prefer to look at the data numerically, preferably in hex. ASCII art now and then, but strictly ironically.

2 Relevant Coursework

1. Which of the following math courses have you taken (i.e. Things that you presumably knew at one point, and could potentially remember with some review):
   - ☐ Linear algebra (matrix algebra, vector spaces, orthogonal matrices, eigenvalues, projections, span)
   - ☐ Linear algebra with proofs
   - ☐ Real analysis
   - ☐ Probability theory (e.g. conditional expectations, law of large numbers, central limit theorem)
   - ☐ Statistics (bias, variance, confidence intervals, basic parametric probability distributions)
   - ☐ Multivariate [differential] calculus (gradients, Jacobians, chain rule)

2. Which of the following topics are you already familiar with from machine learning (they are important machine learning topics that you are assumed to know already for this course):

   - ☐ Supervised learning framework
   - ☐ Cross-validation
   - ☐ Overfitting
   - ☐ Sample bias
   - ☐ Precision/recall, AUC, ROC curves, confusion matrices

3. Other relevant coursework or comments on coursework:

3 Current Knowledge

1. When you hear or see the following, what do you think? (Not whether you already know what's written, but whether you're comfortable with the notation and/or language.)

   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.

   - ☐ You're speaking my language - totally comfortable.
   - ☐ Familiar, but rusty. I'll be ready to go by the start of class.
   - ☐ Never properly learned this. I need to get up to speed.
   - ☐ Wait, this is what I signed up for?

2. When you hear or see the following, what do you think? (Not whether you already know what's written, but whether you're comfortable with the notation and/or language.)

   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 of w.

   ☐ You're speaking my language - totally comfortable.

   ☐ Familiar, but rusty. I'll be ready to go by the start of class.

   ☐ Never properly learned this. I need to get up to speed.

   ☐ Wait, this is what I signed up for?

3. When you hear or see the following, what do you think? (Not whether you already know what's written, but whether you're comfortable with the notation and/or language.):

   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 $\mathcal{X} \times \mathcal{Y}$.

   ☐ You're speaking my language - totally comfortable.

   ☐ Familiar, but rusty. I'll be ready to go by the start of class.

   ☐ Never properly learned this. I need to get up to speed.

   ☐ Wait, this is what I signed up for?

4. When you hear or see the following, what do you think? (Not whether you already know what's written, but whether you're comfortable with the notation and/or language.):

   If we fix a direction $u \in \mathbf{R}^d$, we can compute the directional derivative $f'(x; u)$ as

   $$f'(x; u) = \lim_{h \to 0} \frac{f(x + hu) - f(x)}{h}.$$

   ☐ You're speaking my language - totally comfortable.

   ☐ Familiar, but rusty. I'll be ready to go by the start of class.

   ☐ Never properly learned this. I need to get up to speed.

   ☐ Wait, this is what I signed up for?

5. How comfortable are you answering the following question:

   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

   $$x^T M x - 2b^T x \quad = \quad \left(x - M^{-1}b\right)^T M(x - M^{-1}b) - b^T M^{-1} b \tag{1}$$

   - ☐ So easy. If I had a whiteboard here, I'd do it for you right now.
   - ☐ Yeah - easy. I'll have the answer to you in 5 minutes – I just have to check something on Google first.
   - ☐ Hmmmm. This will be easy by the first day of class.
   - ☐ :(

6. How comfortable are you answering the following question:

   Take the gradient of the following w.r.t. w:

   $$L(w, b, \xi, \alpha, \lambda) = \frac{1}{2}||w||^2 + \frac{c}{n}\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i\left(1 - y_i\left[w^T x_i + b\right] - \xi_i\right) - \sum_{i=1}^{n}\lambda_i\xi_i$$

   - ☐ So easy. If I had a whiteboard here, I'd do it for you right now.
   - ☐ Yeah - easy. I'll have the answer to you in 5 minutes – I just have to check something on Google first.
   - ☐ Hmmmm. This will be easy by the first day of class.
   - ☐ :(

7. How comfortable are you answering the following question:

   Consider $x_1, \ldots, x_n$ sampled i.i.d. from a distribution P on $\mathbf{R}$. Write $\mu = \mathbb{E}x$, for $x \sim P$. Show that the mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an unbiased estimate of μ (i.e. show that $\mathbb{E}\bar{x} = x$). Similarly, show that the sample variance $\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ is an unbiased estimate for $\mathrm{Var}\left(x\right)$.

   - ☐ So easy. If I had a whiteboard here, I'd do it for you right now.
   - ☐ Yeah - easy. I'll have the answer to you in 5 minutes – I just have to check something on Google first.
   - ☐ Hmmmm. This will be easy by the first day of class.
   - ☐ :(